DOCUMENT RESUME

ED 053 325 VT 013 672

AUTHOR	Anderson, Betty R., Ed.; Rogers, Martha P., Ed.
TITLE	Personnel Testing and Equal Employment Opportunity.
INSTITUTION	Equal Employment Opportunity Commission, Washington, D.C.
PUB DATE	Dec 70
NOTE	58p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (74-609942, $.55)
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Employment Opportunities, Equal Opportunities (Jobs), *Guidelines, *Minority Groups, *Personnel Selection, Prognostic Tests, *Testing, Test Interpretation

ABSTRACT

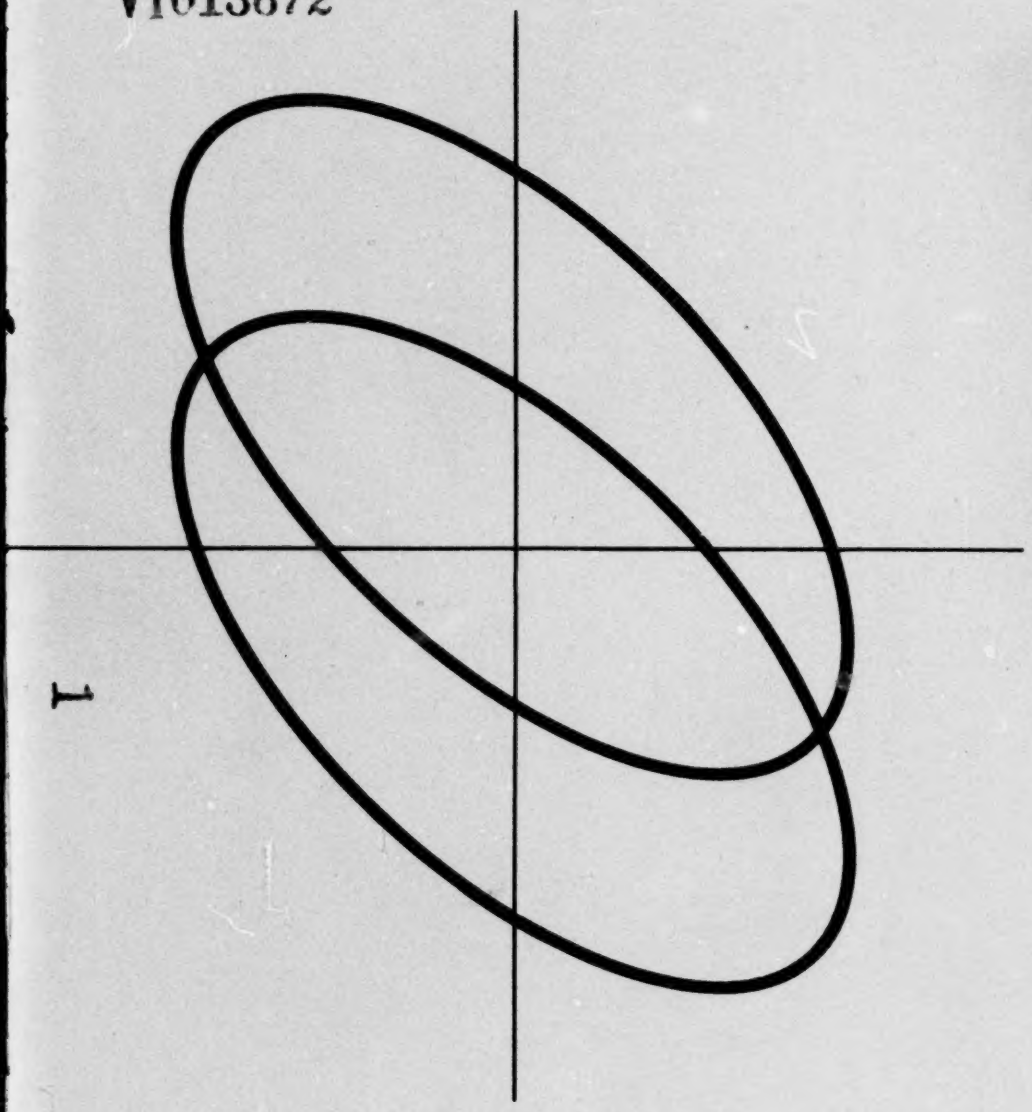
Presented in this book is a series of papers on one of the most complex issues the Equal Employment Opportunities Commission has faced in the first 5 years of its operation. Employment testing represents a systemic problem in equal employment opportunity for minorities and women. Papers included in this document are "Testing of Minority Group Applicants for Employment" by Wallace, Kissinger, and Reynolds, and the following speeches by William Enneis: (1) "Statement Before the House Post Office and Civil Service Subcommittee on Postal Operations," (2) "Discrimination: Planned and Accidental," (3) "Personnel Testing and Equal Employment Opportunity," (4) "Misuses of Tests," (5) "Minority Employment Barriers From the EEOC Viewpoint," (6) "Statement on Personnel Testing and Selection," and (7) "Uses of Nontest Variables in the Government Employment Setting." Also included are guidelines on employee selection procedures. (Author/JS)

VT013672

ED053325

Equal Employment Opportunity Commission

Personnel Testing And Equal Employment Opportunity

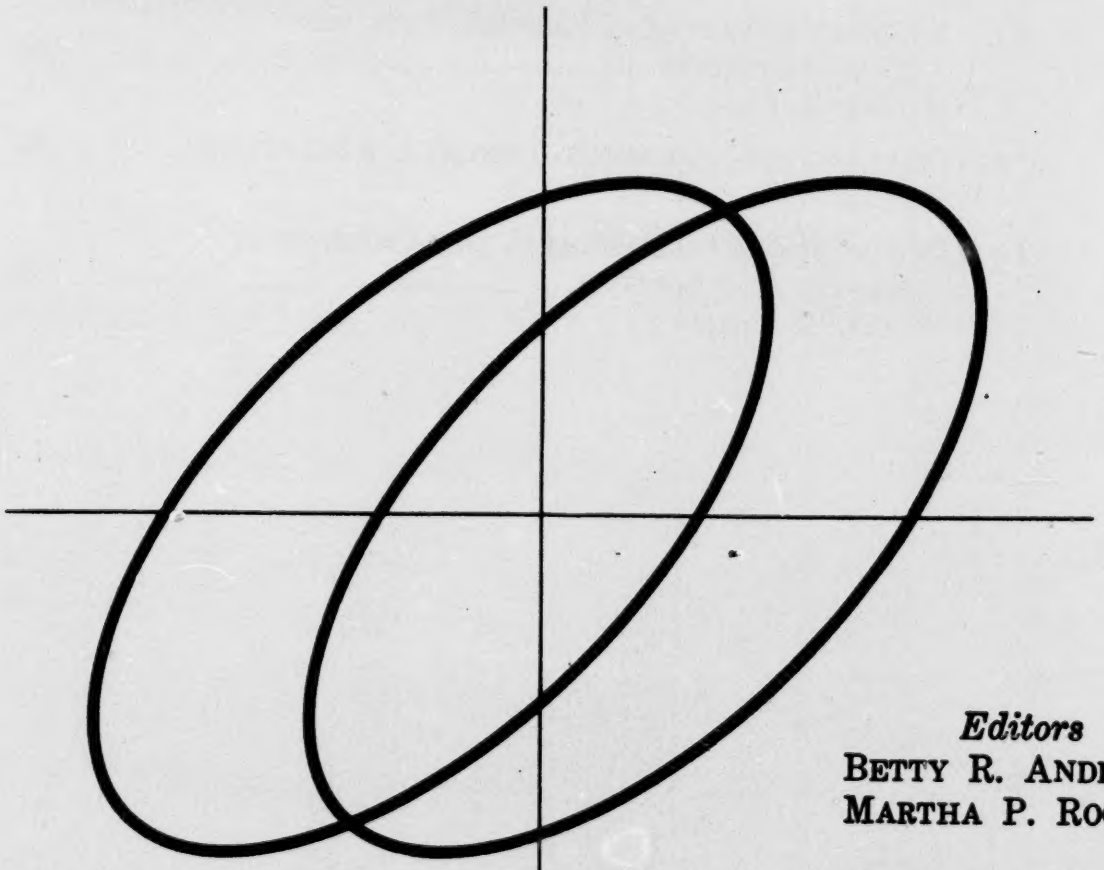


©

Equal Employment Opportunity Commission

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Personnel Testing And Equal Employment Opportunity



Editors
BETTY R. ANDERSON
MARTHA P. ROGERS

Contents

		Page
Foreword		iv
Guidelines on Employee Selection Procedures		v
Chapter		
I.	TESTING OF MINORITY GROUP APPLICANTS FOR EMPLOYMENT Phyllis Wallace Beverly Kissinger Betty Reynolds	1
II.	STATEMENT BEFORE THE HOUSE POST OFFICE AND CIVIL SERVICE SUBCOMMITTEE ON POSTAL OPERATIONS William H. Enneis	16
III.	DISCRIMINATION: PLANNED AND ACCIDENTAL William H. Enneis	21
IV.	PERSONNEL TESTING AND EQUAL EMPLOYMENT OPPORTUNITY William H. Enneis	25
V.	MISUSES OF TESTS William H. Enneis	30
VI.	MINORITY EMPLOYMENT BARRIERS FROM THE EEOC VIEWPOINT William H. Enneis	36
VII.	STATEMENT ON PERSONNEL TESTING AND SELECTION William H. Enneis	42
VIII.	USE OF NONTEST VARIABLES IN THE GOVERNMENT EMPLOYMENT SETTING William H. Enneis	45

December 1970

Library of Congress Catalog Card Number: 74–609942

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402—Price $.55

Foreword

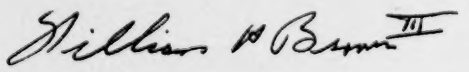
This book presents a series of papers on one of the most complex issues the Commission has faced in the first five years of its operation. Employment testing, like seniority provisions in labor-management agreements, represents a systemic problem in equal employment opportunity for minorities and women. Earlier discriminatory practices, based on now illegal grounds, have often been perpetuated by the use of employment procedures that produce results associated with minority group status and sex but have no demonstrated business necessity.

Thus, on many employment tests minorities suffer from lower scores that bar them from good, productive work while no evidence exists that they will do less well on the jobs for which they were tested. This situation led the Commission to publication of its *Guidelines on Employment Testing Procedures* in 1966 and the subsequent issuance of its *Guidelines on Employee Selection Procedures* in 1970.

The Commission is firmly convinced that equitable selection procedures can provide a large part of the solution to discriminatory employment. However, this equity cannot masquerade under the false front of objectivity. An objective procedure is not necessarily fair, and a fair procedure does not have to be objective.

The fairness of an employee selection procedure resides in its application and the effect its use has on employment opportunities of any person or group protected by Title VII of the Civil Rights Act of 1964. An employment method that screens out a disproportionately high number of minorities or women and has no demonstrated job-relatedness cannot be considered fair, regardless of its structural objectivity or the "good faith" in which it is purportedly used. An unvalidated employee selection procedure does not permit anyone to say that an applicant is either qualified or unqualified to perform satisfactorily on a given job.

I strongly urge you to adopt the idea that merit employment for all Americans means *the ability to do a job*—not just preconceived notions that derive from whim, inadequate knowledge, and outmoded tradition.



WILLIAM H. BROWN, III
Chairman
Equal Employment Opportunity Commission

December 1970
Washington, D.C.

Guidelines on Employee Selection Procedures

FEDERAL REGISTER

VOLUME 35

Saturday, August 1, 1970

PAGE 12333

Title 29—LABOR

Chapter XIV—Equal Employment Opportunity Commission

PART 1607—GUIDELINES ON EMPLOYEE SELECTION PROCEDURES

By virtue of the authority vested in it by section 713 of title VII of the Civil Rights Act of 1964, 42 U.S.C., section 2000e–12, 78 Stat. 265, the Equal Employment Opportunity Commission hereby issues Title 29, Chapter XIV, § 1607 of the Code of Federal Regulations.

These Guidelines on Employee Selection Procedures supersede and enlarge upon the Guidelines on Employment Testing Procedures, issued by the Equal Employment Opportunity Commission on August 24, 1966. Because the material herein is interpretive in nature, the provisions of the Administrative Procedure Act (5 U.S.C. 553) requiring notice of proposed rule making, opportunity for public participation, and delay in effective date are inapplicable. The Guidelines shall be applicable to charges and cases presently pending or hereafter filed with the Commission.

Sec.

1607.1 Statement of purpose.
1607.2 "Test" defined.
1607.3 Discrimination defined.
1607.4 Evidence of validity.
1607.5 Minimum standards for validation.
1607.6 Presentation of validity evidence.
1607.7 Use of other validity studies.
1607.8 Assumption of validity.
1607.9 Continued use of tests.
1607.10 Employment agencies and employment services.
1607.11 Disparate treatment.
1607.12 Retesting.
1607.13 Other selection techniques.
1607.14 Affirmative action.

AUTHORITY: The provisions of this Part 1607 issued under Sec. 713, 78 Stat. 265, 42 U.S.C. sec. 2000e–12.

§ 1607.1 Statement of purpose.

(a) The guidelines in this part are based on the belief that properly validated and standardized employee selection procedures can significantly contribute to the implementation of nondiscriminatory personnel policies, as required by title VII. It is also recognized that professionally developed tests, when used in conjunction with other tools of personnel assessment and complemented by sound programs of job design, may significantly aid in the development and maintenance of an efficient work force and, indeed, aid in the utilization and conservation of human resources generally.

(b) An examination of charges of discrimination filed with the Commission and an evaluation of the results of the Commission's compliance activities has revealed a decided increase in total test usage and a marked increase in doubtful testing practices which, based on our experience, tend to have discriminatory effects. In many cases, persons have come to rely almost exclusively on tests as the basis for making the decision to hire, transfer, promote, grant membership, train, refer or retain, with the result that candidates are selected or rejected on the basis of a single test score. Where tests are so used, minority candidates frequently experience disproportionately high rates of rejection by failing to attain score levels that have been established as minimum standards for qualification.

It has also become clear that in many instances persons are using tests as the basis for employment decisions without evidence that they are valid predictors of employee job performance. Where evidence in support of presumed relationships between test performance and job behavior is lacking, the possibility of discrimination in the application of test results must be recognized. A test lacking demonstrated validity (i.e., having no known significant relationship to job behavior) and yielding lower scores for classes protected by title VII may result in the rejection of many who have necessary qualifications for successful work performance.

(c) The guidelines in this part are designed to serve as a workable set of standards for employers, unions and employment agencies in determining whether their selection procedures conform with the obligations contained in title VII of the Civil Rights Act of 1964. Section 703 of title VII places an affirmative obligation upon employers, labor unions, and employment agencies, as defined in section 701 of the Act, not to discriminate because of race, color, religion, sex, or national origin. Subsection (h) of section 703 allows such persons "* * * to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin."

§ 1607.2 "Test" defined.

For the purpose of the guidelines in this part, the term "test" is defined as any paper-and-pencil or performance measure used as a basis for any employment decision. The guidelines in this part apply, for example, to ability tests which are designed to measure eligibility for hire, transfer, promotion, membership, training, referral or retention. This definition includes, but is not restricted to, measures of general intelligence, mental ability and learning ability; specific intellectual abilities; mechanical, clerical and other aptitudes; dexterity and coordination; knowledge and proficiency; occupational and other interests; and attitudes, personality or temperament. The term "test" includes all formal, scored, quantified or standardized techniques of assessing job suitability including, in addition to the above, specific qualifying or disqualifying personal history or background requirements, specific educational or work history requirements, scored interviews, biographical information blanks, interviewers' rating scales, scored application forms, etc.

§ 1607.3 Discrimination defined.

The use of any test which adversely affects hiring, promotion, transfer or any other employment or membership opportunity of classes protected by title VII constitutes discrimination unless: (a) the test has been validated and evidences a high degree of utility as hereinafter described, and (b) the person giving or acting upon the results of the particular test can demonstrate that alternative suitable hiring, transfer or promotion procedures are unavailable for his use.

§ 1607.4 Evidence of validity.

(a) Each person using tests to select from among candidates for a position or for membership shall have available for inspection evidence that the tests are being used in a manner which does not violate § 1607.3. Such evidence shall be examined for indications of possible discrimination, such as instances of higher rejection rates for minority candidates than nonminority candidates. Furthermore, where technically feasible, a test should be validated for each minority group with which it is used; that is, any differential rejection rates that may exist, based on a test, must be relevant to performance on the jobs in question.

(b) The term "technically feasible" as used in these guidelines means having or obtaining a sufficient number of minority individuals to achieve findings of statistical and practical significance, the opportunity to obtain unbiased job performance criteria, etc. It is the responsibility of the person claiming absence of technical feasibility to positively demonstrate evidence of this absence.

(c) Evidence of a test's validity should consist of empirical data demonstrating that the test is predictive of or significantly correlated with important elements of work behavior which comprise or are relevant to the job or jobs for which candidates are being evaluated.

(1) If job progression structures and seniority provisions are so established that new employees will probably, within a reasonable period of time and in a great majority of cases, progress to a higher level, it may be considered that candidates are being evaluated for jobs at that higher level. However, where job progression is not so nearly automatic, or the time span is such that higher level jobs or employees' potential may be expected to change in significant ways, it shall be considered that candidates are being evaluated for a job at or near the entry level. This point is made to underscore the principle that attainment of or performance at a higher level job is a relevant criterion

in validating employment tests only when there is a high probability that persons employed will in fact attain that higher level job within a reasonable period of time.

(2) Where a test is to be used in different units of a multiunit organization and no significant differences exist between units, jobs, and applicant populations, evidence obtained in one unit may suffice for the others. Similarly, where the validation process requires the collection of data throughout a multiunit organization, evidence of validity specific to each unit may not be required. There may also be instances where evidence of validity is appropriately obtained from more than one company in the same industry. Both in this instance and in the use of data collected throughout a multiunit organization, evidence of validity specific to each unit may not be required: *Provided*, That no significant differences exist between units, jobs, and applicant populations.

§ 1607.5 Minimum standards for validation.

(a) For the purpose of satisfying the requirements of this part, empirical evidence in support of a test's validity must be based on studies employing generally accepted procedures for determining criterion-related validity, such as those described in "Standards for Educational and Psychological Tests and Manuals" published by American Psychological Association, 1200 17th Street NW., Washington, D.C. 20036. Evidence of content or construct validity, as defined in that publication, may also be appropriate where criterion-related validity is not feasible. However, evidence for content or construct validity should be accompanied by sufficient information from job analyses to demonstrate the relevance of the content (in the case of job knowledge or proficiency tests) or the construct (in the case of trait measures). Evidence of content validity alone may be acceptable for well-developed tests that consist of suitable samples of the essential knowledge, skills or behaviors composing the job in question. The types of knowledge, skills or behaviors contemplated here do not include those which can be acquired in a brief orientation to the job.

(b) Although any appropriate validation strategy may be used to develop such empirical evidence, the following minimum standards, as applicable, must be met in the research approach and in the presentation of results which constitute evidence of validity:

(1) Where a validity study is conducted in which tests are administered to applicants, with criterion data collected later, the sample of subjects must be representative of the normal or typical candidate group for the job or jobs in question. This further assumes that the applicant sample is representative of the minority population available for the job or jobs in question in the local labor market. Where a validity study is conducted in which tests are administered to present employees, the sample must be representative of the minority groups currently included in the applicant population. If it is not technically feasible to include minority employees in validation studies conducted on the present work force, the conduct of a validation study without minority candidates does not relieve any person of his subsequent obligation for validation when inclusion of minority candidates becomes technically feasible.

(2) Tests must be administered and scored under controlled and standardized conditions, with proper safeguards to protect the security of test scores and to insure that scores do not enter into any judgments of employee adequacy that are to be used as criterion measures. Copies of tests and test manuals, including instructions for administration, scoring, and interpretation of test results, that are privately developed and/or are not available through normal commercial channels must be included as a part of the validation evidence.

(3) The work behaviors or other criteria of employee adequacy which the test is intended to predict or identify must be fully described; and, additionally, in the case of rating techniques, the appraisal form(s) and instructions to the rater(s) must be included as a part of the validation evidence. Such criteria may include measures other than actual work proficiency, such as training time, supervisory ratings, regularity of attendance and tenure. Whatever criteria are used they must represent major or critical work behaviors as revealed by careful job analyses.

(4) In view of the possibility of bias inherent in subjective evaluations, supervisory rating techniques should be carefully developed, and the ratings should be closely examined for evidence of bias. In addition, minorities might obtain unfairly low performance criterion scores for reasons other than supervisors' prejudice, as, when, as new employees, they have had less opportunity to learn job skills. The general point is that all criteria need to be examined to insure freedom from factors which would unfairly depress the scores of minority groups.

(5) Differential validity. Data must be generated and results separately reported for minority and nonminority groups wherever technically feasible. Where a minority group is sufficiently large to constitute an identifiable factor in the local labor market, but validation data have not been developed and presented separately for that group, evidence of satisfactory validity based on other groups will be regarded as only provisional compliance with these guidelines pending separate validation of the test for the minority group in question. (See § 1607.9). A test which is differentially valid may be used in groups for which it is valid but not for those in which it is not valid. In this regard, where a test is valid for two groups but one group characteristically obtains higher test scores than the other without a corresponding difference in job performance, cutoff scores must be set so as to predict the same probability of job success in both groups.

(c) In assessing the utility of a test the following considerations will be applicable:

(1) The relationship between the test and at least one relevant criterion must be statistically significant. This ordinarily means that the relationship should be sufficiently high as to have a probability of no more than 1 to 20 to have occurred by chance. However, the use of a single test as the sole selection device will be scrutinized closely when that test is valid against only one component of job performance.

(2) In addition to statistical significance, the relationship between the test andcriterion should have practical significance. The magnitude of the relationship needed for practical significance or usefulness is affected by several factors, including:

(i) The larger the proportion of applicants who are hired for or placed on the job, the higher the relationship needs to be in order to be practically useful. Conversely, a relatively low relationship may prove useful when proportionately few job vacancies are available;

(ii) The larger the proportion of applicants who become satisfactory employees when not selected on the basis of the test, the higher the relationship needs to be between the test and a criterion of job success for the test to be practically useful. Conversely, a relatively low relationship may prove useful when proportionately few applicants turn out to be satisfactory;

(iii) The smaller the economic and human risks involved in hiring an unqualified applicant relative to the risks entailed in rejecting a qualified applicant, the greater the relationship needs to be in order to be practically useful. Conversely, a relatively low relationship may prove useful when the former risks are relatively high.

§ 1607.6 Presentation of validity evidence.

The presentation of the results of a validation study must include graphical and statistical representations of the relationships between the test and the criteria, permitting judgments of the test's utility in making predictions of future work behavior. (See § 1607.5(c) concerning assessing utility of a test.) Average scores for all tests and criteria must be reported for all relevant subgroups, including minority and nonminority groups where differential validation is required. Whenever statistical adjustments are made in validity results for less than perfect reliability or for restriction of score range in the test or the criterion, or both, the supporting evidence from the validation study must be presented in detail. Furthermore, for each test that is to be established or continued as an operational employee selection instrument, as a result of the validation study, the minimum acceptable cutoff (passing) score on the test must be reported. It is expected that each operational cutoff score will be reasonable and consistent with normal expectations of proficiency within the work force or group on which the study was conducted.

§ 1607.7 Use of other validity studies.

In cases where the validity of a test cannot be determined pursuant to § 1607.4 and § 1607.5 (e.g., the number of subjects is less than that required for a technically adequate validation study, or an appropriate criterion measure cannot be developed), evidence from validity studies conducted in other organizations, such as that reported in test manuals and professional literature, may be considered acceptable when: (a) The studies pertain to jobs which are comparable (i.e., have basically the same task elements), and (b) there are no major differences in contextual variables or sample composition which are likely to significantly affect validity. Any person citing evidence from other validity studies as evidence of test validity for his own jobs must substantiate in detail job comparability and must demonstrate the absence of contextual or sample differences cited in paragraphs (a) and (b) of this section.

§ 1607.8 Assumption of validity.

(a) Under no circumstances will the general reputation of a test, its author or its publisher, or casual reports of test utility be accepted in lieu of evidence of validity. Specifically ruled out are: assumptions of validity based on test names or descriptive labels; all forms of promotional literature; data bearing on the frequency of a test's usage; testimonial statements of sellers, users, or consultants; and other nonempirical or anecdotal accounts of testing practices or testing outcomes.

(b) Although professional supervision of testing activities may help greatly to insure technically sound and nondiscriminatory test usage, such involvement alone shall not be regarded as constituting satisfactory evidence of test validity.

§ 1607.9 Continued use of tests.

Under certain conditions, a person may be permitted to continue the use of a test which is not at the moment fully supported by the required evidence of validity. If, for example, determination of criterion-related validity in a specific setting is practicable and required but not yet obtained, the use of the test may continue: *Provided:* (a) The person can cite substantial evidence of validity as described in § 1607.7 (a) and (b); and (b) he has in progress validation procedures which are designed to produce, within a reasonable time, the additional data required. It is expected also that the person may have to alter or suspend test cutoff scores so that score ranges broad enough to permit the identification of criterion-related validity will be obtained.

§ 1607.10 Employment agencies and employment services.

(a) An employment service, including private employment agencies, State employment agencies, and the U.S. Training and Employment Service, as defined in section 701(c), shall not make applicant or employee appraisals or referrals based on the results obtained from any psychological test or other selection standard not validated in accordance with these guidelines.

(b) An employment agency or service which is requested by an employer or union to devise a testing program is required to follow the standards for test validation as set forth in these guidelines. An employment service is not relieved of its obligation herein because the test user did not request such validation or has requested the use of some lesser standard than is provided in these guidelines.

(c) Where an employment agency or service is requested only to administer a testing program which has been elsewhere devised, the employment agency or service shall request evidence of validation, as described in the guidelines in this part, before it administers the testing program and/or makes referral pursuant to the test results. The employment agency must furnish on request such evidence of validation. An employment agency or service will be expected to refuse to administer a test where the employer or union does not supply satisfactory evidence of validation. Reliance by the test user on the reputation of the test, its author, or the name of the test shall not be deemed sufficient evidence of validity (see § 1607.8(a)). An employment agency or service may administer a testing program where the evidence of validity comports with the standards provided in § 1607.7.

§ 1607.11 Disparate treatment.

The principle of disparate or unequal treatment must be distinguished from the concepts of test validation. A test or other employee selection standard—even though validated against job performance in accordance with the guidelines in this part—cannot be imposed upon any individual or class protected by title VII where other employees, applicants or members have not been subjected to that standard. Disparate treatment, for example, occurs where members of a minority or sex group have been denied the same employment, promotion, transfer or membership opportunities as have been made available to other employees or applicants. Those employees or applicants who have been denied equal treatment, because of prior discriminatory practices or policies, must at least be afforded the same opportunities as had existed for other employees or applicants during the period of discrimination. Thus, no new test or other employee selection standard can be imposed upon a class of individuals protected by title VII who, but for prior discrimination, would have been granted the opportunity to qualify under less stringent selection standards previously in force.

§ 1607.12 Retesting.

Employers, unions, and employment agencies should provide an opportunity for retesting and reconsideration to earlier "failure" candidates who have availed themselves of more training or experience. In particular, if any applicant or employee during the course of an interview or other employment procedure claims more education or experience, that individual should be retested.

§ 1607.13 Other selection techniques.

Selection techniques other than tests, as defined in § 1607.2, may be improperly used so as to have the effect of discriminating against minority groups. Such techniques include, but are not restricted to, unscored or casual interviews and unscored application forms. Where there are data suggesting employment discrimination, the person may be called upon to present evidence concerning the validity of his unscored procedures as well as of any tests which may be used, the evidence of validity being of the same types referred to in §§ 1607.4 and 1607.5. Data suggesting the possibility of discrimination exist, for example, when there are differential rates of applicant rejection from various minority and nonminority or sex groups for the same job or group of jobs or when there are disproportionate representations of minority and nonminority or sex groups among present employees in different types of jobs. If the person is unable or unwilling to perform such validation studies, he has the option of adjusting employment procedures so as to eliminate the conditions suggestive of employment discrimination.

§ 1607.14 Affirmative action.

Nothing in these guidelines shall be interpreted as diminishing a person's obligation under both title VII and Executive Order 11246 as amended by Executive Order 11375 to undertake affirmative action to ensure that applicants or employees are treated without regard to race, color, religion, sex, or national origin. Specifically, the use of tests which have been validated pursuant to these guidelines does not relieve employers, unions or employment agencies of their obligations to take positive action in affording employment and training to members of classes protected by title VII.

The guidelines in this part are effective upon publication in the FEDERAL REGISTER.

Signed at Washington, D.C., 21st day of July 1970.

[SEAL] WILLIAM H. BROWN III,
Chairman.

[F.R. Doc. 70-9962; Filed, July 31, 1970; 8:46 a.m.]

CHAPTER I

Testing of Minority Group Applicants for Employment

PHYLLIS WALLACE
BEVERLY KISSINGER
BETTY REYNOLDS*

March 1966

This report is the result of intensive research on a highly controversial and complex subject. A number of psychologists have provided us with data from their current studies. We are especially grateful to Commissioner Hernandez who permitted us to examine the testing materials from her files. Dr. Robert Krug, Director of Research for the Peace Corps, made available several of his studies on testing of minority persons. Dr. Philip Ash, Research Assistant to the Vice President of Industrial and Public Relations for Inland Steel; Dr. Joel Campbell of Educational Testing Service; Mr. Howard C. Lockwood of the Lockheed Aircraft Corporation; Dr. Richard Shore, Policy Planning Staff of the Department of Labor; and Dr. Arthur Brayfield, Executive Secretary of the American Psychological Association, have sent us a number of articles. While acknowledging our debt to various scholars, we, of course, assume full responsibility for any errors of fact or interpretation.

* * *

INTRODUCTION

The Motorola case** and the Tower amendment to Title VII of the Civil Rights Act of 1964 Section 703(h)*** have dramatized the issue of whether the use of general intelligence tests by employers as selection devices for hiring and promotion deprives Negroes and members of other minority groups of equal employment opportunity. Individuals from culturally disadvantaged**** backgrounds perform less well on these tests on the average than do applicants from middle class environments and consequently may be screened out of training programs and/or excluded from jobs. Differences in culture, in opportunity, and in experience can have a devastating effect on test performance. Since many Negroes, Mexican-Americans, Indians, and lower-class whites have not shared the middle class culture, they may perform in an inferior manner on tests of general intelligence, particularly paper and pencil, but not necessarily on per-

*Now Betty R. Anderson

**See Appendix A for the chronology of the Motorola case.

***". . . nor shall it be an unlawful employment practice for an employer to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin." Sec. 703(h)

****See Appendix B.

formance for which the tests are supposed to be predictive.

Consistent and significant differences on mean scores are also found between age, sex, educational, and urban-rural groups, but the focus of this report is the effect of testing on the culturally disadvantaged, many of whom are Negroes. This report is not concerned with the willful misuse of tests to discriminate such as giving tests to Negroes but not to whites, or requiring Negroes to achieve higher scores than whites, or failing Negroes regardless of their actual performance. These practices are clearly unlawful. The question to be considered here is whether many "professionally developed ability tests" used by employers to select qualified applicants do in fact discriminate *inadvertently*.

Authorities in the field of psychological testing have suggested several proposals for mitigating the effects of *unintentional* types of discrimination against minority groups. We have examined the various proposals and have concluded that careful selection and administering of tests and validation of the testing instrument within an industrial setting, may be the most desirable means to achieve the goal of full utilization of the nation's human resources. The implications of this affirmative conclusion are discussed from the viewpoint of the Equal Employment Opportunity Commission, private employers, and the research psychologists who would have to assume the major responsibility for formulating suitable standards for selection of testing programs.

TYPES OF TESTS

The major types of tests most commonly used in employee selection are: (1) general intelligence tests, (2) tests of specific intellectual abilities, (3) knowledge and skill tests, (4) measures of dexterity and coordination, and (5) inventories of personality traits.

Intelligence tests such as the Wonderlic, Stanford-Binet, and Otis Quick-Scoring are designed primarily to measure the ability of the individual to understand and to reason with words and numbers. Such tests are most useful in selection for jobs where learning from and understanding verbal academic material is important.

Specific intellectual abilities tests determine potential for learning certain kinds of work and for solving certain kinds of problems. The tests are not designed to test for a specific job, but to measure the skills for understanding and reasoning with words, numbers and symbols, visualizing of spatial relationships, word fluency, visual speed and accuracy, and creative abilities.

Knowledge and skill tests are usually specific to a job or job family. Knowledge tests are designed to measure the understanding of blueprint reading, electronics, accounting, etc., while skill tests measure one's ability to type, to take dictation, to drive, etc. These tests measure the degree or level of knowledge or skill already attained by candidates at the time of the test.

Dexterity and coordination tests measure speed and accuracy of physical movements. These tests must be very specific to the movements required in the job and are usually constructed by the employer. Examples of such tests are spatial and mechanical abilities, perceptual accuracy, motor abilities.

Personality and interest tests are intended to indicate how a person typically acts and feels, and to determine the type of activities he likes. Tests of this nature have been developed primarily for use in either vocational guidance or clinical use. *It is extremely important for a highly trained professional psychologist to evaluate and interpret the results of these tests.*

Tests may be further categorized as ap-

titude versus proficiency. Aptitude tests are designed to measure potential while achievement tests measure skill level at the time of testing.

HOW TESTS DISCRIMINATE AGAINST MINORITY GROUPS

An aptitude test that fails to predict job performance in the same way for both Negroes and whites, or fails to predict job performance at all is not a valid test. If such a test is weighted to differentiate between Negroes and whites, it is similarly invalid and similarly discriminatory. Tests may be held to discriminate in the *social sense* if they deny equal opportunity for consideration. A test may operate in this manner (a) when scores on it tend to differentiate between identifiable sub-groups, where the sub-grouping itself is not a relevant selection factor, and either (b) scores for the lower group underpredict performance on the job when the standards of the upper-group are applied, or (c) scores on the test do not predict job performance for either group.[1]

It is known that Negroes on the average do less well on paper and pencil tests than whites. The mean scores for Negroes are lower than the mean scores* for whites on most paper and pencil tests of general ability, intelligence, aptitude, learning ability, or overall ability. The distribution of scores overlap, often considerably, but the average scores differ significantly in most studies.

More research has been done on the testing of minority group children than workers, but the information which has resulted from this research offers insight into why Negro adults achieve a lower mean score than job applicants from a more middle class background. Newton S. Metfessel, psychologist at the University of Southern California, in his research on children and youth who live in the culture of poverty, found that cultural factors such as home and family structure, personality and social characteristics, learning characteristics, and general school relationships handicap performance on tests.

These children usually come from a home environment where there is such a paucity of objects that the child's conceptual formation development is adversely affected. They also lack curiosity, and this affects both motivational patterns and the development of creative behavior. The culturally disadvantaged child is characterized by weak ego-development, a lack of self-confidence, and a negative self-concept. These conflicting feelings about himself frequently result in exaggerated positive and negative attitudes towards others.[2]

Many aspects of learning characteristics are affected by the culturally poor background. The culturally disadvantaged typically have a cognitive style which responds more to visual and kinesthetic signals than to oral or written stimuli. Also, these children learn more readily by inductive than deductive approaches. Learning experiences which move from the part to the whole rather than from the whole to the part are invariably more successful. Significant gaps in knowledge and uneven patterns of learning are typical of this type of background.

Children from the culture of poverty have had little experience in receiving approval for success in learning a task, an assumption on which the school culture is organized.

> The cycle of skill mastery which demands that successful experiences generate more motivation to perform which in turn guarantees levels of skill sufficient to prevent discouragement, and so on, may be easily reversed in direction and end the achievement habit prior to its beginning.[3]

*Raw scores are converted to norms in order to compare an individual performance with a specific group. See glossary in Appendix B.

In general school relationships and school characteristics, these children from the background of cultural deprivation are placed at a marked disadvantage on standardized tests, which for the most part have been designed to test the white, middle class child. The shortcomings of the standardized tests when they are used with disadvantaged minority groups are discussed below.

Reliability of Differentiation

Standardized tests may not provide reliable differentiation in the range of the minority group's scores. The reliability coefficient for a particular test is strongly affected by the spread of test scores in the group for which the reliability is established. In general, the greater the spread of scores in the reliability samples, the higher the reliability coefficient. For many tests, there is evidence

> that children from the lower socio-economic levels tend to have a smaller spread of scores than do children from middle-income families, and such restriction in the distribution of scores tends to lower reliability so far as differentiation of measurement with such groups is concerned.[4]

Predictive Validity

Second, the predictive validity of tests for minority groups may be quite different from that for the standardization and validation groups. Factors which may impair a test's predictive validity are:

1. **Test-related factors**, i.e., test taking skills, anxiety, motivation, speed, understanding of test instructions, degree of item or format novelty, examiner-examinee rapport which may affect test scores but have little relation to the criterion.

2. **Complexity of criteria**—It is important to recognize the influence of other factors, not measured by tests, which may contribute to criterion success. Since disadvantaged groups tend to do poorly on general intelligence and achievement tests of the paper and pencil type, one should explore background, personality, and motivation of members of such groups for compensatory factors, untapped by the test, which may be related to criterion performance.[5]

While certain aptitude and proficiency tests may have excellent criterion validity for some purposes, even the best of them are unlikely to reflect the true capacity of underprivileged children. They tap abilities that have been molded by the cultural setting. The test content, mode of communication involved in responding to test items, and the motivation needed for making responses are intrinsically dependent upon the cultural context.[6]

Validity of Test Interpretation

Third, the validity of the interpretation of tests is strongly dependent upon an adequate understanding of the social and cultural background of the group in question. Sources of error in test interpretation stemming from lack of recognition of the special features of culturally disadvantaged groups are: (1) deviation error—tendency to infer maladjustment from responses which are deviant from the viewpoint of a majority culture, but which may be typical of a minority group; (2) simple determinant error—thinking of the test content as reflecting some absolute or pure trait, process, factor, or construct, irrespective of conditions of measurement or the population being studied; (3) failure barriers—requiring minority group individuals to solve problems with unfamiliar tools.[7]

Job applicants from lower socio-economic levels may be characterized in con-

trast to their middle class counterparts as being less verbal, more fearful of strangers, less confident, less motivated toward scholastic and academic achievement, less conforming to middle class norms of behavior and conduct, less knowledgeable about the world outside their immediate neighborhood. To the extent that these sub-cultural differences affect test performance adversely, these persons may be denied the opportunity to employment and a more productive contribution to society. Selection instruments often call for responses that are influenced by the culture of the applicant's community or quality of his educational opportunity. Since such tests are "culturally loaded" against persons from a lower socio-economic status, they may operate as instruments of racial discrimination. The crucial question is whether employers use techniques that *unwittingly* eliminate persons who might perform satisfactorily on the job. The relationship between test performance and cultural deprivation on the one hand, and job performance on the other, must be investigated for both white and nonwhite job applicants.

PROPOSED SOLUTIONS TO THE PROBLEM OF CULTURAL BIAS IN TESTING

Most employers defend tests as an efficient device for choosing the most qualified applicants. Where Negro job applicants consistently score significantly below white job applicants, a question should be raised about test scores as predictors of job performance. In an employment situation we would like to know whether differences between group means are also associated with performance on the criterion. Do the factors that depress test performance also depress trainability or whatever criterion is to be predicted? Psychologists have suggested ways in which the effect of cultural bias inherent in many aptitude tests can be alleviated for minority group applicants. Few of these proposals have been universally accepted, but most have been discussed in the professional literature on testing of minority groups and the culturally disadvantaged.

Variants of "Culture-Free" and "Culture-Fair" Tests

1. **Culture-free tests**—One such proposal is the development of tests which are free of cultural bias in their content and instructions. Dr. Robert Krug, who has written extensively on testing of minority persons, indicates that one of two conditions must be met before a test can be classified as "culture-free": either the test items are those which all people of all cultures have had equal opportunity and equal motive to learn, or the test items must possess complete novelty for all people of all cultures.[8] For all practical purposes these two conditions are almost impossible to meet and the idea is often rejected as unfeasible. Howard Lockwood of Lockheed Corporation states that many industrial psychologists agree that even if such a test could be developed, it would be useless in personnel selection. It is impossible, he maintains, to avoid measuring cultural influences, and if they were completely eliminated from all tests, the tests would measure, in essence, nothing.[9]

2. **Culture-fair tests**—Dr. Krug, on the other hand, does not reject the idea entirely. He describes a "*culture-fair*" test, as a modification of the "*culture-free*" idea. The assumption underlying the "*culture-fair*" tests is that there exists a set of test stimuli which are equally appropriate, that is, equal opportunity and motive to learn, for at least two cultural groups.[10] Dr. Paul Schwartz, who headed an AID-sponsored aptitude test development project in West Africa, has done most of the research in this area. A "culture-fair" test

or "culture-common test" developed by Schwartz for Nigerian and American children utilized a set of fruits and vegetables which were approximately equal in familiarity to both cultures.

3. **Culture-equivalent tests** — Dr. Schwartz also developed another variant of this concept called "*cultural-equivalent*" tests, denoting that two tests which are not identical may, in fact, be equivalent. In this case investigations were undertaken to discover cultural counterparts of tools and machines, cultural manifestations of mechanical principles, and cultural opportunities to acquire information of potential relevance to mechanical training.[11] The argument of cultural equivalence rests on the demonstration that tests constructed in this way have been valid predictors of performance in Westernized training programs in shop mechanics, electrical repair, and the like. Development of similar tests in this country is impeded by lack of knowledge concerning the culture of southern Negroes, northern slum-dwellers of all races, or any other identifiable sub-groups. Dr. Ash asserts that so-called culture-fair tests do not measure aptitudes or characteristics significantly related to most ordinary measures of job success such as turnover, production or foreman ratings.[12]

Creativity Tests

Another approach, adopted by Dr. Newton Metfessel and Professor J. J. Risser, of the University of Southern California, involves the use of tests to measure *creativity* rather than traditional intelligence tests. The latter sample only a relatively small portion of the factors which are involved in intellectual potential and have placed a premium on verbal comprehension and speed of response and emphasize convergent thinking, or the ability to select the one correct answer.[13]

Creativity tests, on the other hand, stress divergent thinking or the ability to create new or original answers. They are, according to Metfessel, more suitable for the testing of the culturally disadvantaged and certain ethnic groups whose command of language is not highly developed.

These tests utilize the most common and familiar of objects in order to sample the testee's ability to recognize problems, and his originality, flexibility, and fluency of thinking. Tasks include suggesting improvements in a familiar device such as a telephone, or thinking of problems that might occur in the use of an object such as a candle. One test requires the subject to list as many uses as he can for a broom handle.[14]

The tests are scored simply on the number of acceptable answers given by the subject. They seem to be as effective in predicting academic success as traditional intelligence tests and, probably, would be as effective as the latter in predicting job performance.

Differential Selection among Applicants from Different Socio-Economic Ethnic Backgrounds

It has been proposed that, since prediction equations for job performance for most tests currently in use have been based on the performance of whites, different standards (separate test norms, conversion tables, prediction weights, etc.) be employed for Negroes and other culturally disadvantaged groups. This approach involves a technique known as the *moderator* variable. Applicants for a given job are divided into subgroups, and selection procedures are applied differentially to members of the two groups. Applicants could be classified, for example, on the basis of a measure of socio-economic status, demographic data (such as percentage

of Negroes living in the census tract from which the applicant is applying), and race.

Studies could then be undertaken to determine whether there is, in fact, a difference in the predictive efficiency of job tests as between high and low status groups. Differences in selection procedures for different ethnic groups do not mean a lowering of standards because the standards which count are standards of performance on the job, not the selection standards. Equally qualified pesons may be selected from various ethnic groups by applying the standards which are appropriate to each group.[15]

Lockwood has proposed the use of "cultural exposure" as a moderator variable. Examinees should be grouped homogeneously as to cultural exposure and these groups treated separately in validity studies. Cultural exposure is defined as the material things to which a person has been exposed and the attitudes to which he has been exposed and which he has acquired. Research would lead to a better identification of the culturally disadvantaged and to the utilization of their abilities through a refinement in prediction of training and occupational success.[16]

A major investigation is under way by Dr. Richard Barrett to determine if the division of applicants into sub-groups improves the accuracy of prediction for members of both groups. If selection is improved by applying different procedures to the high and low socio-economic groups, then the more talented would benefit, regardless of race.

It may also happen that dividing the group of applicants on the basis of race may lead to improved accuracy of predictions for members of both races. Such a result has far reaching implications for fair employment practices because *failure to treat the two races separately would, if current policies were followed, lead to discrimination against the more talented Negroes.*[17]

The overwhelming evidence is that the cultural background of the Negro in America today is so different from that of the white that his performance during the selection process can reasonably be expected to be different. It may be difficult to find an adequate sample of Negroes in most occupations in order to develop separate and suitable prediction equations for them. Lockwood also cautions against the use of a lower minimum score or separate standards of test performance for Negroes since it might tend to perpetuate the idea of race differences or race inferiority.

Dual Test Standards and Compensatory Training

The concept of a "dual standard" has some support among psychologists. Ash cites the work of Dr. Kenneth B. Clark of the City University of New York. Clark's work suggests that culturally deprived people who score low on tests may tend to overachieve on the job. In studying the college performance of students who scored low on college entrance tests, Clark found that for students from nondeprived environments, the tests were good predictors, and low college entrance test scores were accurate indicators of poor grades. On the other hand, students coming from deprived environments did significantly better in college than would have been predicted from the tests.[18]

An experimental training program run by the Federal Department Stores in Detroit, Michigan, indicates that a lowering of required test scores will not necessarily result in a lower quality of job performance. The Federal Department Stores took 16 young people from culturally and economically deprived areas, all of whom had failed standard employment tests and were classified as "unemployable," and put them through a 10-week special training

program. All 16 subsequently were employed, 14 at Federal and two elsewhere. The record of performance of all 14 employees at Federal exceeded what was predicted by standard sales aptitude tests. Some exceeded the company's minimum performance standards for new employees by "unbelievable margins."[19]*

Although the Federal Department Stores experiment is considered one of the first of its kind in offering compensatory training for individuals with low test scores, the concept of "double-standard" has had wide acceptance for years in the fairly common practice of maintaining different norms for the sexes. Several popular tests which offer different sex norms are the Bennet, the Wonderlic, the Minnesota Paper Form Board, and the Thurstone Temperament Schedule.

It is generally agreed that some of these sex differences on tests are undoubtedly of environmental origin. Girls are expected to score lower than boys on tests of mechanical information. It is also expected that girls will perform less effectively on tasks for which the Mechanical Information test is a predictor. This, however, does not prevent many companies from employing women in manufacturing tasks which require mechanical ability where they perform satisfactorily.[20]

On the basis of these examples, it appears that a "double-standard" can be justified in some circumstances, though a double standard in job performance and hiring of less qualified applicants is usually rejected as not being effective. If it can be demonstrated that score X for Group A and Score K-k for Group B are associated with *identical* levels of performance on the job, then an employer might reasonably consider adopting a more flexible attitude toward test scores.[21]

*Re-test results one year later for the ten trainees still employed by Federal showed no significant changes in the scores as a group.

Intensification of Recruitment

While there are significant differences in average performance, there is a considerable overlap in the distribution of test scores of whites and Negroes. It has been proposed, on the basis of this observation, that employers who wish to maintain their present standard of performance on their pre-employment tests can increase their number of Negro employees by intensifying recruitment among Negroes in order to identify those whose test performance is equal to that of acceptable white applicants. Although this approach has merit in that it could provide employment for Negroes who are qualified but who do not apply for jobs in companies where they assume discrimination is practiced, it is not a solution to the testing problem. It ducks the question of the fairness of tests to those who fail because of cultural disadvantage, and it will not provide enough additional workers to satisfy present and future labor needs.[22]

Use of Test Scores as Only One Indicator

One other practical solution similar in many respects to the "double-standard" is to use test scores as only one indicator among others in the hiring decision, with a clear awareness that, where the applicant has not shared in the predominant middle-class verbal culture, the test score significantly underestimates his potential. A difference of one point more or less cannot be expected to determine if an applicant will fail or succeed on the job. Other personal characteristics such as achievement, motivation, and dependability may be just as significant indicators of successful job performance, and they usually can be identified in each cultural group.

Proper Testing Practices

Along with adoption of a more flexible attitude toward test scores, the most immediate improvement can be accomplished by an emphasis by the employer on proper testing practices.

1. The employer could reconsider the relevance of the qualifications for employment to the specific job tasks required by his company. Many of these requirements are stated in terms of some generalized stereotypes, such as high school graduate, high IQ, or potential to advance to higher level jobs, and are quite extraneous to the requirements of that job. Tests should be professionally chosen to fit the distinctive features of both the industry and the background, education, and other characteristics of the successful work force. It is unreasonable to insist that all lower level workers have potential for supervisory positions. An employer may eventually find that by adopting a more reasonable set of qualifications for each job, he will have access to a considerably larger source of workers who can perform capably and who will present him with fewer problems of employee frustration or labor turnover.

2. Selection tests should be developed by reputable professional psychologists who are competent in conducting testing programs in an industrial setting.

3. Pre-employment tests should be administered by personnel who are properly trained not only in the technical details of giving tests, but also in the orientation and handling of people in the testing situation. Members of disadvantaged groups tend to be particularly sensitive to any mannerisms that might be considered antagonistic, sarcastic, or condescending, and test administrators should be aware of this and be able by their behavior to alleviate a certain amount of test anxiety. A personnel manager at a recent testing conference complained that the number of Negro applicants for jobs in his company had fallen off by 80 percent after the company recently instituted a pre-employment testing program.*

4. A policy of re-testing "failure" candidates may gain for an employer many good employees who otherwise would have been eliminated by the first test. Many candidates, particularly members of minority groups, regard testing as a threatening situation and do not perform as well as they could. A second test would provide a more accurate indication of the true capability of a person who is less experienced with testing situations and who may have been intimidated by his first experience.

5. Finally, the most important principle is *validation* of tests in order to confirm the relationship between test scores and on-the-job performance. There is general agreement that tests should not be used for a group which differs from the validation group. Validity is relative both to the criterion to be predicted and to the group for which the prediction is to be made. Very few employers have validated their testing instruments. In a recent survey by the University of Wisconsin Industrial Relations Research Center, 152 companies which apply testing techniques were canvassed and only 7 percent reported that all their tests had been validated locally against on-the-job performance measures. Nearly 60 percent had validated *none* of their tests. The remainder reported that some but not all of their tests were validated.[23]

Dr. Warren Ketcham, University of Michigan psychologist and Vice President of Psychodynamics Research and Associates, has suggested that within company norms should be used exclusively. This only requires that an applicant perform on

*University of Michigan Testing of Minority Group Applicants, January 26, 1966.

tests as well as or better than persons who have done or are presently doing the job satisfactorily. The norm tables should then be used to rank applicants as sub-standard, low-average, average, high-average, or superior.[24]

From recent discussions with research psychologists attached to large industrial concerns, it appears that many companies are developing ability tests which will measure the essentials required for training or employment, while keeping at a minimum the relevant aspects of culture. For a number of reasons, these findings may never be released for general consumption. One of the responsibilities of the Commission will be to encourage this type of research by the psychological profession. If the purpose of tests is to uncover talent and potential, irrespective of label, surely the Commission could not advocate a more commendable policy.

UNITED STATES AS A MODEL EMPLOYER

If the Equal Employment Opportunity Commission establishes basic guidelines on testing of minority group applicants, including a provision on validation of tests, it will require private employers to satisfy certain standards which the United States government, as a civilian employer, for the most part does not meet.*

The U.S. government has set a fine example in its standardized testing program for the military where these tests have been completely validated. Testing in the Armed Forces serves a number of major programs, two of which are (1) to identify the number of personnel required in each skill and professional category and (2) to identify each individual for training, upgrading, and utilization to his highest potential.

In order to maintain validity, test development activities are mainly serviced by professional job analysts, subject matter specialists, and test psychologists and validated in the working area. This systematic approach is essential to assure that the tests sample specific job functions in direct proportion to the importance of those functions to the job. As a result, job analysis provides not only a basis for test construction, selection and training, but also a means for increasing productivity and facilitating work.

RECOMMENDATIONS FOR TESTING GUIDELINES

The following recommendations are designed as a guide to help employers establish objective standards for selection, screening, and promotion of workers. These procedures should ensure that all qualified applicants are given equal opportunity for employment.

1. Job descriptions should be examined and their *critical* requirements established before tests are selected for screening applicants.

2. Tests used should be those developed by reputable psychologists. Such tests should be administered by professionally qualified personnel who have had training in occupational testing in an industrial setting.

3. Rigidly inflexible minimum scores should be re-examined in light of the considerable research under way on differential selection.

4. Test scores must be considered as only one source of information, and must be combined with other available data on performance such as motivation, leader-

*Of some interest is the fact that the United States Employment Service has recently undertaken a program to develop aptitude measures that can be used to evaluate potential for literacy training, vocational training and occupational potential of the educationally deficient. Much of the research is designed to improve the General Aptitude Test Battery (GATB).[25]

ship and organizational experience, self-sufficiency, and dependability.

5. Tests should be validated within the setting where they will be used. Validation should be for as many separate groups as possible in preference to one large heterogeneous group.

6. It may be advisable for employers who deal with applicants from culturally disadvantaged backgrounds to offer retests to candidates who are unsuccessful on their first try, since these people are less familiar with testing situations and may not perform as well as they are able.

APPENDIX A

CHRONOLOGY OF THE MOTOROLA CASE

1. *July 15, 1963.*—Leon Myart, a Negro, applied for a job as a television phaser and analyzer at the Franklin Park plant of Motorola, Inc. Myart took a five minute intelligence test (General Ability Test No. 10), was interviewed, and was sent home without being told whether he qualified for employment.

2. *July 29, 1963.*—Failing to receive a job offer, Myart filed a complaint with the Illinois Fair Employment Practices Commission and the President's Committee on Equal Employment Opportunity alleging that his not being hired was due to racial discrimination.

3. *January 27–28, 1964.*—Hearing of the Motorola case before hearing examiner Robert Bryant of the Illinois Fair Employment Practices Commission.

4. *February 26, 1964.*—The hearing examiner directed that Myart be offered a job, that test No. 10 should no longer be used, and that any new test developed in its place should "reflect and equate inequalities and environmental factors among the disadvantaged and culturally deprived groups." He argued that the test had been normed on "advantaged groups" and did not "lend itself to equal employment opportunity to qualify for the hitherto culturally deprived and disadvantaged groups'"

5. *April 18, May 25, July 14–15, 1964.*—Review of the Motorola case before the full Commission.

6. *November 18, 1964.*—The Commission issued its unanimous decision, finding that Myart had been denied employment because of his race and while not supporting the order to hire Myart directed that he be compensated one thousand dollars.

7. *April 27, 1965.*—Illinois Circuit Court decision on appeal of Motorola. The ruling requiring Motorola to pay Myart one thousand dollars was reversed, but the Commission's findings on discrimination were upheld.

8. *November 11, 1965.*—Case argued before the Illinois Supreme Court.

9. *March 24, 1966.*—Illinois Supreme Court reversed the judgment of the circuit court on grounds that the alleged unfair employment practice was not established by a preponderance of the evidence.

APPENDIX B

GLOSSARY OF SPECIAL TERMS

Criterion—A standard that provides a basis for evaluating the validity of a test.

Cultural bias—Propensity of a test to reflect favorable or unfavorable effects of certain types of cultural backgrounds.

Culture-fair test—A test yielding results that are not culturally biased.

Culture-free test—A test yielding results that are not influenced in any way by cultural background factors.

Norms—Statistics that depict the test performance of specific groups. Grade, age, and percentile are the most common types of norms.

Reliability—The degree of consistency, stability, or dependability of measurement afforded by a test.

Validity—The extent to which a test measures the trait for which it is designed, or for which it is being used, rather than some other trait.

Psychological test—An observation of a sample of human behavior made under standard, controlled conditions which results in a linear evaluation called a score.

Culturally disadvantaged—Groups which do not have full participation in American society because of low incomes, substandard housing, poor education, and other "atypical" environmental experiences.

APPENDIX C

SELECTED REFERENCES ON TESTING

1. American Psychological Association, Committee on Scientific and Professional Responsibility, "Social Influences on the Standards of Psychologists," *American Psychologist*, Vol. 19, 1964, pp. 167–173.

2. *American Psychologist, Special Issue: Testing and Public Policy*, American Psychological Association, Vol. 20, No. 11, November, 1965.

3. Ash, Philip, "Fair Employment Practices Commission Experiences with Psychological Testing," *American Psychologist*, September 1965, pp. 747–798.

4. Ash, Philip, "Race, Employment Tests, and Equal Opportunity." (Presented before Conference of National Association of Inter-Group Relations Officers, Chicago, Illinois, October 21, 1965.)

5. Ash, Philip, "The Implications of the Civil Rights Act of 1964 for Psychological Assessment in Industry." (Presented as part of a symposium, "Legal Issues Which Confront the Psychologist and the Community," 72 Annual APA Convention, Chicago, Illinois, September 5, 1965.)

6. Barrett, Richard S., "Differential Selection Among Applicants from Different Socio-Economic Ethnic Backgrounds," *Selecting and Training Negroes for Managerial Positions*, Princeton, New Jersey, Educational Testing Service, 1965, pp. 91–100.

7. Campbell, Joel, "Testing of Culturally Different Groups," *Research Bulletin*, Princeton, New Jersey, Educational Testing Service, No. RB 64–34, June, 1964.

8. "Can Today's 'Unemployables' Become Tomorrow's Salesmen." (Reprinted with permission from McGraw-Hill, Inc.), New York, New York, American Jewish Committee, March 29, 1965.

9. Chambers, Yolande, "Retraining Program Upsets Test Predictions," *Personnel Service*, September-October, 1965.

10. Clark, Kenneth B., "Color, Class, Personality, and Juvenile Delinquency," *Journal of Negro Education*, Vol. 28, 1959, pp. 240–251.

11. Coles, Robert, *The Desegregation of Southern Schools: A Psychiatric Study*, New York, New York, Anti-Defamation League of B'nai B'rith, 1963.

12. Culhane, Margaret M., "Testing the Disadvantaged," *The Journal of Social Issues*, April, 1964.

13. Dreger, Ralph M., and Miller, Kent S., "Recent Research in Psychological Comparisons of Negroes, and Whites in the United States." (Presented at Southeastern Psychological Association, Atlanta, Ga., April 2, 1965.)

14. Dvorak, Beatrice, *et al.*, "New Directions in U.S. Employment Service Aptitude Test Research," *Personnel and Guidance Journal*, October 1965.

15. Fandell, Todd E., "Testing and Discrimination," *Wall Street Journal*, April 21, 1965.

16. French, Robert L., "The Motorola Case," *The Industrial Psychologist APA Newsletter*, Division of Industrial Psychology of the American Psychological Association, Vol. 2, No. 3, August, 1965.

17. Ghiselli, E. E., "A Summary of the Validities of Occupational Aptitude Tests." (Presented before the Western Psychological Association, 1965.)

18. Ghiselli, E. E., "Differentiation of Tests in Terms of the Accuracy with which They Predict for a Given Individual," *Educational Psychological Measurement*, Vol. 20, 1960, pp. 675–684.

19. Goslin, D.A., *The Search for Ability: Standardized Testing on Social Perspective*, New York: Russell Sage Foundation, 1963.

20. "Guidelines for Testing Minority Group Children." (Prepared by a Work Group of the Society for the Psychological Study of Social Issues, Division 9 of the American Psychological Association.) *Journal of Social Issues Supplement*, Vol. XX, November 2, 1964.

21. Guion, Richard, "Subjectivity in Hiring Standards," *Personnel Hiring*, McGraw-Hill, 1965, pp. 490–493.

22. Katz, I., "Review of Evidence Relating to Effects of Desegregation on the Intellectual Performance of Negroes," *American Psychologist*, Vol. 19, 1964, pp. 381–399.

23. Ketcham, Warren, "Testing Minority Group Applicants." (Prepared for the University of Michigan Bureau of Industrial Relations Personnel Techniques Seminars, January 26, 1966.)

24. Klineberg, Otto, "Negro-White Differences in Intelligence Test Performance: A New Look at an Old Problem," *American Psychologist*, Vol. 18, 1963, pp. 198–203.

25. Krug, Robert E., "Some Suggested Approaches for Test Development and Measurement." (Presented at the symposium, "The Industrial Psychologist, Selection and Equal Employment Opportunity," 1964 Convention of APA, Los Angeles, California, September 4–9, 1964.)

26. Krug, Robert E., "The Problem of Cultural Bias in Selection," *Selecting and Training Negroes for Managerial Positions*, Princeton, New Jersey, Educational Testing Service, 1965.

27. Laney, A. R., "Scientific Hiring of Appliance Servicemen," *American Gas Association Monthly*, January, 1951.

28. Lockwood, Howard C., "Critical Problems in Achieving Equal Employment." (Presented at symposium, "The Industrial Psychologist, Selection and Equal Employment Opportunity," American Psychological Association 1964 Convention, Los Angeles, California, September, 1964.)

29. Lockwood, Howard C., "Cultural Exposure and Race as Variables in Predicting Training and Job Success."

30. Lockwood, Howard C., "Testing Minority Applicants for Employment." (Presented at 1964 Annual Convention of the California State Psychological Association.) *Personnel Journal*, Vol. 44, July-August 1965, pp. 356–360.

31. Lockwood, Howard C., "Progress in Plans for Progress for Negro Managers." (Presented at the Executive Study Conference, New York, New York, November 10, 1964.)

32. Metfessel, Newton S., "Conclusions from Previous Research Findings Which Were Validated by the Research and Evaluation Conducted by the Staff of Project Potential," University of Southern California, 1965.

33. Metfessel, Newton S., and Risser, J. J., "Project Potential: Interpretive Guide for the Tests of Creativity," 1965.

34. Ricklefs, Roger, "Jobs and Psychology: Personnel Tests Win Widening Business Use," *Wall Street Journal*, February 1965.

35. Rosenberg, Leon A., Rosenberg, Anna M., and Stroud, Michael, "The Johns Hopkins Perceptual Test (The Development of a Rapid Intelligence Test for the Pre-School Child)," April, 1966.

36. Runney, George, "Enforcement of Fair Employment Under Civil Rights Act of 1964," *University of Chicago Law Review*, Vol. 32, 1965, pp. 430–470.

37. Scioto, Leonard A., and Ryterband, Edward, "Civil Rights and the Industrial Psychologist: A Challenge Not a Threat," *The Industrial Psychologist*, Vol. 2, 1965, pp. 40–43.

38. Smith, Karl, "Civil Rights and Psychological Testing," *Experimental Cybernetic Foundations of Learning Science*, Madison, Wisconsin, University of Wisconsin.

39. Smith, Karl, "Cybernetic Analysis of Personality Assessment Procedures," and "Cybernetic Analysis of Psychological Testing and Test Prediction," *Experimental Behavioral Cybnernetics*, Madison, Wisconsin, University of Wisconsin, June 4, 1965.

40. Smith, Karl, "Proposal for a National Institute of Work Science," *Experimental Cybernetic Foundations of Learning Science*, Madison, Wisconsin, University of Wisconsin, 1963.

41. *Selecting and Training Negroes for Managerial Positions*, Educational Testing Service, Princeton, New Jersey, November 1965.

42. Spock, Benjamin, "Children and Discrimination." (Reprinted from *Redbook*), American Jewish Committee, New York, February 1965.

43. Tumin, Melvin M. (Editor), *Race and Intelligence*, Anti-Defamation League of B'nai B'rith, New York, 1963.

Motorola Case

44. Circuit Court of Cook County, Illinois
Motorola, Inc. vs. Illinois Fair Employment Practices Commission and Leon Myart (Report of Proceedings)

45. In the Matter of
Leon Myart and Motorola, Inc., State of

Illinois, Fair Employment Practices Commission Charge No. 63C–127

46. Supreme Court of Illinois
Motorola, Inc. vs. Illinois FEPC and Leon Myart (Brief of Plaintiff-Appellant)

47. Supreme Court of Illinois
Motorola, Inc. vs. Illinois FEPC and Leon Myart (Reply Brief of Plaintiff-Appellant)

48. Supreme Court of Illinois
Motorola, Inc. vs. Illinois Fair Employment Practices Commission and Leon Myart
(Brief and Argument for Illinois Fair Employment Practices Commission, Defendant-Appellee)

49. Supreme Court of Illinois, September Term, A.D. 1965
Motorola, Inc. vs. Illinois Fair Employment Practices Commission and Leon Myart
(Appeal from the Circuit Court)

APPENDIX D

SOURCES*

1. Ash, Philip (5), p. 9.
2. Metfessel, Newton (32), p. 3.
3. *Ibid.*, p. 4.
4. Guidelines (20), p. 131.
5. *Ibid.*, p. 136.
6. *Ibid.*, p. 137.
7. *Ibid.*, p. 139–142.
8. Krug, Robert (25), p. 6.
9. Lockwood, Howard (30), p. 4.
10. Krug (25), p. 7.
11. *Ibid.*, p. 8.
12. Ash, Philip (4), p. 11.
13. Metfessel, Newton (33), p. 1.
14. *Ibid.*, p. 3.
15. Ash (4), p. 13.
16. Lockwood (29), p. 4.
17. Selecting and Training Negroes for Managerial Positions (41), p. 93.
18. Ash (4), p. 12.
19. *Merchandising Week* (8).
Chambers, Yolande (9).
20. Ash (4), p. 5.
21. Krug (25), p. 6.
22. Ash (4), p. 1.
23. *Ibid.*, (4), p. 4.
24. Ketcham (23), p. 3.

*Numbers in parentheses refer to Appendix C.

CHAPTER II

Statement Before the House Post Office and Civil Service Subcommittee[1]

WILLIAM H. ENNEIS

Mr. Chairman, members of the Committee, I am glad to appear before you today to explain some of the issues that confront our society in the areas of employment testing and assessment of the qualifications of employees for advancement to more responsible positions in their chosen fields of work. The ever increasing complexity of business and industrial activities, with rapid introduction of mechanically complex labor-saving devices, has led to efforts that will increase the effectiveness of present workers and improve the quality of new employees. At the management end of these enterprises, change is possibly even greater—not so much in what must be done but more in the concepts and techniques that must be understood and used for some reasonable maintenance or improvement of a competitive position. For those organizations that refuse to follow the need for rapid change, both in technology and personnel procedures, recent economic history is filled with examples of companies, even entire industries, that have sunk into a morass of inefficiency, from which a recovery is increasingly difficult, if not impossible.

Today, I am going to talk about the efforts that the American business and industrial sector must undertake if we are to avoid the enormous waste of our national human resources by application of inappropriate personnel selection and promotion methods. More specifically, I shall discuss the use of psychological tests. However, what I say is equally applicable to other personnel assessment methods that are used on job applicants and present employees.

The history of tests goes back to the 1890's when they were largely experimental, laboratory-type devices that were mostly scientific curiosities. These early tests were often based on physiological, perceptual, and motor activities and generally failed to predict academic achievement, for which purpose they were designed. At that time little thought and effort was directed to the prediction of successful performance among industrial workers.

It was not until 1905 that Alfred Binet and Theodore Simon constructed the first successful intelligence test. They had been commissioned by the City of Paris to produce a test that would predict which pup-

[1] U. S., Congress, House, Committee on Post Office and Civil Service, *Personnel Promotion System of the Post Office Department*, before a subcommittee of the Committee on Post Office and Civil Service, House of Representatives, 90th Cong., 1st sess., 1967, pp. 44-64.

ils were most likely to require special instruction to remain in school. In 1917, A. S. Otis developed a paper-and-pencil intelligence test that could be administered to large groups of people at one time. During World War I, the Army Alpha and the Army Beta tests were used to classify recruits. The Beta test was nonverbal and designed for soldiers who could neither read nor write. Following World War I, both the Army Alpha (a verbal, paper-and-pencil test) and the Otis, in several varieties and revisions, were used to predict academic learning from the elementary grade school levels through college entrance.

I include this very brief and incomplete history of tests because it is well to keep in mind that most of the present tests for employee selection retain the highly verbal, academic flavor that characterized the Otis and the Army Alpha. Indeed, today one of the most popular employment tests is the Wonderlic Personnel Test (in several different forms), which is essentially a shortened version of the older Otis tests (also published in several forms).

The notion still persists among many employers and some psychologists that general intelligence—as measured by tests with heavy emphasis on verbal ability, numerical ability, some aspects of spatial ability, and abstract reasoning—is a prerequisite for satisfactory job performance. However, many years of research, involving results from hundreds of studies on different kind of workers, fail to support this assumption.

Dr. Edwin E. Ghiselli, an eminent industrial psychologist, has summarized the usefulness of general intelligence tests for predicting performance on various types of jobs. He says that general intelligence tests are virtually worthless in estimating the job performance of computing clerks, service workers, mechanical repairmen, machine workers, and sales clerks. However, intelligence tests have demonstrated probable usefulness in selection of managers, inspectors, and general clerks. Even in these latter jobs the average relationship between test scores and job performance is low enough that only a small part of an employee's job performance can be attributed to his relative position within the group of persons who take the test.

The main point to be made here is that general intelligence tests are not highly predictive of many types of work in business and industry. And why are they not predictive? Because the content of general intelligence tests is not related, by and large, to what people are required to do as workers on those jobs.

At this point, I would like to introduce the *Guidelines on Employment Testing Procedures*, issued by the Equal Employment Opportunity Commission last year. These guidelines were produced in response to many questions and issues generated by the part of Section 703(h), in Title VII of the Civil Rights Act of 1964, that specifically allows an employer

> . . . to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin.

The EEOC *Guidelines* are introduced because they embody the substance of good personnel employment practices as recommended by experts in industrial psychology and personnel administration for the past forty to fifty years. For example, the Commission advocates that careful job analyses be conducted to determine the essential requirements of the job before any tests are chosen. Indeed, the job analyses may clearly indicate that no tests whatsoever are called for in selecting some categories of employees.

I shall mention that I have encountered situations where a general intelligence test

is used to screen applicants for the positions of janitor, dishwasher, window washer, and laborer. This practice borders on the absurd, especially when persons in these normally dead-end jobs are not in any lines of promotion where advancement to higher positions occurs on a seniority basis.

The real tragedy lies not in the employer's waste of his money in the use of irrelevant selection devices, but rather in the fact that minority group applicants and older people tend to be excluded from employment, whereas otherwise they might have been selected on the basis of more meaningful requirements.

The reasons that minority group applicants, on the average, earn lower scores than non-minority group applicants are many and complex. However, fewer educational, social, and cultural opportunities are among the more important. Also, minority group applicants probably have had fewer chances to learn to take the many varieties of tests that appear on the American educational and vocational scene.

At this time, I wish to introduce two research reports that deal with tests and minority group applicants. The first, *Testing of Minority Group Applicants for Employment,* is a report produced by the Equal Employment Opportunity Commission. The second document, *The Berkeley Project,* contains the results of a study carried out on minority and non-minority applicants for jobs at the municipal level in California. I especially recommend the first document for a more detailed discussion of the subject than is feasible for me to present here.

With respect to the effect of age on test scores, I shall confine my remarks to the statement that older persons, even those under 40 years, suffer in two major ways in a test situation—other factors being equal, such as natural ability and previous education. First, they have been out of school for several years and have grown unaccustomed to the often extreme time pressures that are brought to bear in the administration of some tests. Second, they have become "rusty," so to speak, on many of the academically oriented items that characterize many preemployment and promotion tests. Yet, older persons may do quite as well on the job as their juniors, and in many instances are better risks because they tend to be more stable and less likely to quit for a position paying a few cents more per hour.

In general—and now I shall speak for myself, but as a psychologist—it is unethical to use tests to determine suitability for promotion of present employees, unless the nature of the test is such that it can clearly demonstrate that some workers would be a danger to either themselves or those around them or that they would be definitely incapable of performing the job to which they might be promoted. An employee's work history is a more reasonable indicator of probable success in a higher job. After all, it is fairly well known that high school grades are the best predictor, by and large, of college grades. There is no reason to believe that performance on the present job is an inherently less suitable predictor of success on a higher job than some test which has less relevance in terms of actual work to be carried out.

There may be, naturally, some instances in which job descriptions, based on job analyses, will indicate that a test, or battery of tests should be used to select initial applicants for jobs. Again it is important to remember that the superficial appearance or so-called "logical" content of a test is not enough to determine its suitability. Many tests that appear "reasonable" in subject matter turn out to be very poor selection instruments.

Insofar as the professional, technical judgments related to the suitability of tests are concerned, I shall submit as evidence the *Standards for Educational and*

Psychological Tests and Manuals, published by the American Psychological Association. Although these standards contain many technical concepts and terms and were written mostly for psychologists and educators, their essential points are not so esoteric as to exceed the potential grasp of any personnel administrator who truly wishes to understand and carry out his responsibilities.

Some applicants will, of course, fail the employer's test(s). The Equal Employment Opportunity Commission advocates the retesting of those persons who do not meet the minimum scores on the first or subsequent test administrations. This recommendation applies to all persons who fail the tests. It does not apply just to members of minority groups.

The EEOC *Guidelines* do not carry any mention of the length of time between retests, but a period of six months is reasonable. If it becomes apparent at any time that the applicant's test failure was due to some unusual circumstance that might be expected to lower anyone's score, the retest should be permitted as soon as the testee's situation has returned to normal. It should be noted that the guidelines refer to ". . . those 'failure candidates' who have availed themselves of more training or experience." In this respect, higher scores on retesting may be the product of that same training or experience, which, in turn, may be reflected in job performance.

The capstone to this entire discussion is whether test scores are related to the adequacy of an employee's work. All other factors aside, the true value of any selection procedure—be it test, interview, personal history, or background data—rests in the ability of that device to predict an applicant's job performance. This relationship between test scores and standards of job performance is called the *validity* of the test. This type of validity is obtained after carefully controlled research and adequate statistical comparison of test and criterion scores. Some tests have zero validity for certain kinds of work, which means that there is no systematic relationship between employees' test scores and their job performance.

The predictive value of an employee selection device is not established by statements that "the test works because I (we) have had years of success in its use." Such pronouncements, without more evidence, are generally made in the absence of a systematic effort to demonstrate the worth of the particular selection procedure(s) in question.

The Equal Employment Opportunity Commission considers test validation such an important matter that its first decision based on the testing *Guidelines* addressed itself to the problems of (1) test validation and (2) disproportionate test failure rates among present Negro employees, who were required to take tests to move into promotion lines which had, before the Civil Rights Act of 1964, not been open to them. This decision, as published by the Bureau of National Affairs in January, 1967, is submitted as a part of my statement today.

The essential points of the decision, related to employment testing, are these. If a test screens out a disproportionate percentage of minority group applicants and has no demonstrated relationship to job performance, then that test:

1. acts as a discriminatory vehicle, regardless of employer intent and
2. is not "professionally developed" within the meaning of Title VII.

Unfortunately, most companies using tests do not validate them. A Prentice-Hall survey (*New Ideas;* October 4, 1966; p. 861), which I am introducing as evidence, showed that only about 50 percent of private employers have taken this step to determine their usefulness. The results of other surveys have generally shown less than 50 percent of employers validating their tests.

A disturbing point in the recent attempt of some employers to get on the bandwagon of testing is that they justify their action by stating that new, complex, labor-saving equipment requires greater skills than were necessary in older production methods. Yet, at the same time, these employers do not adduce objective evidence to show that greater employee skill or judgment is required in operation of the new equipment. In fact, many labor-saving methods probably require lesser amounts of skill, although perhaps of a different type, than the methods which were supplanted.

Within the past decade, some psychologists have begun to recommend differential validity studies on minority groups versus the so-called majority group. The essence of this movement is the realization that various selection procedures, including tests, may not predict the job success of Negroes, for example, in the same way that they predict the success for whites. I wish to introduce into the record the report of a study conducted by Dr. Felix Lopez of the Port of New York Authority. The results are truly amazing, and if differential validities, as reported there, are borne out by future research, the implications for personnel psychology are enormous. With the continuation of differential validity research, more questions may be raised than are settled about the universal application of tests to minorities and non-minorities on the same basis.

Are there any immediate solutions to the problems raised by testing and the issue of equal employment opportunity? Should we, for example, eliminate testing altogether and rely on other personnel assessment methods?

The answer to the last question is "No." The elimination of tests would create more problems than it would settle.

The answer to the first question is "Yes." Employers can use all of the information available to them about applicants instead of relying solely or very heavily on test scores. As an example, minimum cut-off scores should not be rigidly applied without consideration of an applicant's other qualifications. The Equal Employment Opportunity Commission recommends a *total personnel assessment* system that utilizes all available, relevant data on both applicants and present employees. In this approach, no single method or device automatically excludes an individual from further employment consideration.

Another constructive approach to employment testing is differential validation research, which I cited earlier. However, we should go beyond the now topical variables of race, or color, or national origin and explore what it is about our society that produces inter-group differences not only in test scores but also in the way that these scores predict—or fail to predict—job performance. I am certain that the significant variables will not be those based on race, color, or national origin.

If the personnel selection and placement issues are tackled creatively, I believe that this country will go a long way toward solutions to the equal employment problem and the equitable utilization of our national human talents. Our human resources are very possibly our most important asset, and they should not be wasted because of any reticence in attacking the problem of discrimination.

I thank you for the opportunity to appear before your Committee today.

CHAPTER III

Discrimination: Planned and Accidental

WILLIAM H. ENNEIS

Presented March 20, 1967, as part of a symposium, *Psychological Tests and the Law,* at the annual American Personnel and Guidance Association Convention, Dallas, Texas.

* * *

Title VII of the Civil Rights Act of 1964 addresses itself to unlawful employment practices by private employers, employment agencies, and labor organizations. These unlawful practices are ones that would adversely affect an individual's hiring, classification, or promotion because of the person's race, color, religion, sex, or national origin. On the other hand, preferential treatment is expressly forbidden on the basis of existing differentials between employment and census rates. This means that previous inequities of employment cannot be solved legally by Federal requirements for compensatory hiring and promotion of minority group members.

The Equal Employment Opportunity Commission was established under Title VII. As a part of its activities, the Commission is responsible for educational and affirmative action programs with private employers, investigation and conciliation of charges, and technical studies which will advance the purposes of Title VII.

Today, my presentation will be concerned mostly with the use of psychological tests in personnel assessment. The center of this discussion is contained in Section 703(h) of Title VII and reads as follows.

> Nor shall it be an unlawful employment practice for an employer to give and to act upon the results of any professionally developed ability test provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex, or national origin.

The above language is often referred to as the "Tower Amendment" because it was introduced by Senator Tower of Texas during Senate debate on the 1964 Civil Rights Act. Its introduction was prompted by the testing issue raised in the Illinois FEPC case of *Myart vs. Motorola,* which in early 1964 received nationwide attention among employers after the Illinois FEPC hearing examiner ruled, among other things, that the particular employment test did not afford equal employment opportunity to culturally disadvantaged groups and was "obsolete" because the test was standardized on "advantaged" groups. Senator Tower's amendment had the effect of establishing, in legal terms, the right of an employer to give "professionally developed ability tests" as long as they were not intentionally discriminatory.

The word "ability" is emphasized since there may arise a very serious question as to what constitutes an ability test. Certainly most of us here are sure that we

know what an ability test is. Even if we cannot immediately define the term, we are most confident that we could say whether or not Text X is a test of some ability, or abilities—or whether it belongs to another category of psychological instruments, such as the personality, character, attitude, interest, or temperament scales.

The ability test should be looked upon as a *limit response* test, in which each person is essentially asked to perform within the limit of his ability as operationally defined by the item content, format, and time constraints (if any) imposed upon him. This definition of an ability test covers all categories of intelligence, aptitude, achievement, and performance tests. The definition, as I have stated it, should not be confused or associated with the term *power test*, which refers to an ability test administered without formal and strictly applied time limits. Furthermore, the term *limit response* test does not imply any measurement, or attempt at measurement, of capacity.

On the other hand, the scales associated with what we often lump into the category of personality measurement devices can be called *interpretive response* tests. They measure what an individual is willing to say about his usual behavior and his perception and interpretation of the events in his environment. The personnel administrator, educator, or psychologist is not trying to measure what a person *can* do when he uses the *interpretive response* measuring instrument. Therefore, I maintain, strictly as a psychologist, that such *interpretive response* devices do not qualify as ability tests within the meaning of Title VII.

However, the legal interpretation of ability tests may be quite different from the one I have just given. For that matter, the courts, if the occasion ever arises, may *not* differentiate between ability and the so-called personality tests. The reasons for this are at least twofold.

First, it may be decided that the intent of the wording was to allow the use of all tests, the usual psychological and educational definitions notwithstanding. My opinion is that such an approach to the interpretation of a non-legal term, namely ability test, reflects either a gross presumption of psychological expertise or an abysmal failure to distinguish technical and scientific definitions from the lay use of words. It is very important to remember that the English language, including technical vocabulary, is adequate for the expression of any idea in legal form. Therefore, if words are used that detract from or contravene some legislative intent, who is to say what the original thought was if the technical vocabulary of the statute, as enacted, implies another intent? The defect, in this case, would lie not in the wording of the Tower Amendment but the interpretation thereof.

A second interpretation that might be given to the term ability test is one that embraces the concept of empirical validity. After all, if a test can predict the job performance of present or future employees, does it not predict *ability* on the job, regardless of whether or not it is a limit response test? This idea has been expressed by Philip Ash (1966). His thoughts were based on an earlier expression by Miller, Duffy, and Haught (1964).

This second approach to an interpretation arising from litigation over the use of employment tests is more logical and certainly less dogmatic than the question of legislative intent. However, it avoids the main issue of predictor content by shifting responsibility onto the relationship between the test and criterion scores.

Transfer of the burden from the test in this way would virtually demand that all tests be validated by each employer before they could be used in the employment

process. Furthermore, mere validation might not be enough because then the matter of professional development of the criterion would surely arise. So great is the attention to tests that almost no thought is given to the ways in which criterion content and use may give rise to spurious validity coefficients, especially when the validation group contains both minority and non-minority applicants or employees.

The subject of test validation is an appropriate one for broaching the topic of accidental discrimination. My discussion today will not cover such matters as complete and total reliance on test scores, unrealistically high standards for applicants, subjective biases that may occur in the interview, rigid application of test cutting scores, failure to consider an applicant's previous work, and the lack of test validity data for the type(s) of work involved. These matters have been covered by other writers. They are also discussed in the EEOC *Guidelines on Employment Testing Procedures.* It is now time to extend our thought to what can happen when tests are validated without adequate control over extraneous variables, particularly those which influence criteria of job performance.

Let us turn to criterion contamination, which we have all learned is a bad thing in any validity study. Briefly, criterion contamination occurs when employees' performance—or supervisors' evaluation of performance—contains a systematic error unrelated to actual job performance. If this constant error is correlated with test scores, the problem is quite serious in that statistically significant validity coefficients may be obtained, when in fact no true correlation exists between various tests and criteria of job performance. This condition is very likely to occur if, for example, (1) Negro and white employees are present in the same group on which correlation coefficients are computed and (2) supervisors, who are required to complete merit rating forms, have access to employees' test scores in their personnel folders.

Criterion contamination of this type should always be avoided, but it is particularly important to guard against it in the situation described above because it is well known that minority group applicants make, on the average, lower scores than those from the so-called majority group. If supervisors use test scores to assign, and ultimately justify, merit ratings it is easy to see how inflated validity coefficients can occur. And in all due honesty, the employer may believe that he has a perfectly acceptable selection device—even though application of the obtained regression equation on whites and Negros alike will result in disproportionate rejection rates among future Negro applicants. Krug (1966) has presented this type of situation pictorially in terms of theoretical distributions of predictor and criterion scores, although he did not direct his remarks to criterion contamination as such.

It is far preferable, therefore, if at all possible, to validate tests separately on minority and non-minority groups. A study by Lopez (1966) has pointed up the urgency for differential validation research. He showed that certain predictor variables are not correlated in the same way with the job performance of whites and Negroes.

Another pitfall in validation of tests with minority groups is the poorly constructed performance evaluation scale that contains vague names of job behaviors and the level at which they are executed. All of us are familiar with merit rating forms which list the names, but no definitions, of perhaps a dozen traits with scale divisions of "Excellent," "Very Good," "Above Average," "Satisfactory," and "Below Average."

It is all too apparent that spurious validity coefficients for employment tests will occur if supervisors assign most of the

"Excellent" and "Very Good" ratings to the white employees and most of the "Above Average" and "Satisfactory" ratings to the Negroes or Spanish-speaking employees. Note that it would be most difficult and, in all fairness, improper to accuse the supervisors of prejudice because they have, after all, rated most of the minority group employees "Satisfactory" or above. The problem lies not with the supervisors, but rather with the measure(s) of employee performance. The resolution of this issue lies in the careful definition of performance characteristics and behavioral anchoring of levels within traits. We may need a reemphasis of critical incidents and forced-choice scales to provide differentiation among meaningful job related behaviors.

Finally, one must approach the unpleasant fact that some employers do use tests to discriminate, in the legal sense of intentionally, against certain minority groups. This discrimination is usually subtle, however, and no longer takes the blatant form it did prior to 1964. Furthermore, most discrimination in employment is a *local* matter. Rarely, if ever, does a large, national employer practice discrimination everywhere in the country. Indeed, the national, corporate levels of management are, as a collective whole, against discrimination in employment but sometimes allow it to continue on a regional or local level out of sheer inertia—not because of an obdurate resistance to equal employment opportunity.

Proof of the use of tests as the planned vehicle of discrimination is not easy, and the evidence is often circumstantial. I have not encountered any situations where it was apparent that time limits were rigged, different tests were used for minority groups than for the non-minority applicants, or falsification of test records occurred.[1] Sometimes recruiting procedures, percentage of minorities in the employer's work force and labor suppply, distribution of minorities at various levels of work skill, use of other employee selection procedures, promotion practices, and efforts to validate tests are indicators of whether the tests themselves have been chosen in such a way as to screen out a large proportion of minority group applicants.

The EEOC *Guidelines on Employment Testing Procedures* place great emphasis on a total personnel assessment system. These *Guidelines* contain recommendations which, if carefully followed, can help an employer go a long way toward equal employment opportunity.

[1] Since writing this paper, the author has encountered several instances of imposition of unequal, more rigorous requirements on minority applicants.

REFERENCES

Ash, Philip, The implications of the Civil Rights Act of 1964 for psychological assessment in industry. *American Psychologist*, 1966, *21*, 797–803.

Krug, R. E., Some suggested approaches for test development and measurement. *Personnel Psychology, 19*, No. 1, 24–35.

Lopez, Felix, Jr., Current problems in test performance of job applicants: I. *Personnel Psychology, 19*, No. 1, 10–18.

Miller, L H., Duffy, E. R., and Haught, F. B., *Civil Rights Act of 1964*. Washington, D.C.: National Association of Manufacturers, Law Department, July, 1964.

CHAPTER IV

Personnel Testing and Equal Employment Opportunity

WILLIAM H. ENNEIS

Presented June 7, 1969, at the annual meeting of the Pennsylvania Psychological Association, Mt. Pacono, Pennsylvania.

* * *

During the four years that the Equal Employment Opportunity Commission has been in operation, the issues of employment testing have been among the more persistent and difficult ones. As far as the EEOC is concerned, it is probably correct to say that psychological testing involves the most direct confrontation of a scientific discipline with legal definitions of employment discrimination. The purpose of this paper is an explanation of how the Commission approaches the resolution of such matters. To this end, I shall also discuss what I consider to be the simplistic use of certain terms by both psychologists and laymen.

In each Title VII case where employment methods are alleged or believed to affect hiring, promotion, transfer, rate of pay, or any other condition of private employment by virtue of race, color, religion, sex, or national origin, the Commission's policy is to determine first what effect the method has on the group of persons affected under the charge filed. For example, if the charge alleges employment discrimination on the basis of race and the charging party is a Negro, the relative effect of the personnel procedure on Negroes and other employees (or applicants) will be determined. This determination is made from all available and relevant information.

Whenever the class of persons represented by a charging party is adversely affected by the personnel assessment method, the second step is to find out whether that method has some justifiable business function. An example of adverse effect is the use of a test that results in the employment of 60 percent of nonminority applicants but limits employment of the minority group to only 20 percent of all such applicants. When the use of an employment procedure shows no disparity between minority and nonminority groups, the employer's justification of the procedure is not usually a critical point. However, the Commission is not unmindful of the fact that the potential for discrimination exists even if minority group applicants earn the same average test score as nonminorities (French, 1965; Kirkpatrick, Ewen, Barrett, & Katzell, 1968, p. 6).

The justifiable business function of any personnel assessment procedure ultimately boils down to the question as to whether it demonstrably improves the effectiveness of the employer's work force. This improvement can be, and most often is, expressed in the significant results of a validation study. Validation research can be executed with any measurable predictor and criterion, and the Commission empha-

sized the importance of criterion-related validity in its *Guidelines on Employment Testing Procedures* (EEOC, 1966).

Obviously, technical deficiencies in the predictor, the criterion, or the control of extraneous variables can militate against the demonstration of validity; but these matters are the responsibility of the employer, not that of the EEOC. On the other hand, failure to control or adjust for factors that operate to produce spuriously significant results will be questioned.

From the brief foregoing explanation of the way in which the EEOC treats cases involving assessment of worker potential, it can be seen that the Commission's main concern is the *use* of personnel assessment methods. The consideration of intent and design of these methods is often secondary, although imposition of differential hiring standards or opportunities raises the question of both intent and design. Examples of such differential treatment (rather than effect, as explained earlier) are (1) invocation of a maximum weight limit on Negro female applicants when none is imposed on white females, (2) refusal to hire Spanish Surnamed applicants with more than a certain number of dependents while not imposing that restriction on Anglos, and (3) testing Negro applicants only for the specific job(s) which they name on the application blank while testing whites for all vacant jobs in the employer's facility.

In an effort to avoid the differential effect problem (and hence skirt the validity issue?), some psychologists and lay users of our jargon have recommended so-called culture-free tests for employment purposes. Presumably, the "culture-free" test is one that eliminates or considerably reduces inter-ethnic differences in mean test scores. Therefore, each racial, color, etc. group should in the long run have equal opportunities for employment because very nearly equal mean test scores should result in equal proportions of each ethnic group being hired or promoted when the higher scoring persons are always selected, regardless of group affiliation.

However much the reduction of mean test score differences may be touted, most psychologists seem to forget that equalization of variances and skewness among ethnic groups would also be required for real equality of opportunity to exist on the basis of test score alone. Consider an employer with an extremely favorable (in these days!) selection ratio of 0.10. If the standard deviation of test scores is quite a bit larger for the nonminority than for the minority group, assuming a normal distribution for each group, there is a very good likelihood that use of a purportedly "culture-free" test with equal ethnic means will still result in a disproportionately high rejection rate of minority applicants, simply because the higher scoring nonminorities will fill up the selection ratio quota of 10 percent of all applicants before an equal proportion of minorities is hired. With any selection ratio greater than 0.50, however, the minority group would have a higher proportionate representation among those hired—again assuming equal test means of minority and nonminority groups and normal distributions of scores in both.

If the distribution of scores in one ethnic group is highly positively skewed and in the other is highly negatively skewed, the group with negative skewness will be much favored in employment opportunity even though both group means are equal. Finally, from the industrial psychologist's viewpoint one must remember that tests do not and should not always constitute the only factor in personnel actions. Therefore, if raw test scores are added to raw scores from other personnel assessment methods, the test will weight itself in proportion to its standard deviation and the standard deviation of all

other components that make up the composite or total assessment score (Tiffin & McCormick, pp. 527–532). Whenever there are significant differences in the standard deviations among ethnic groups on the test and on other factors determining employability, the picture rapidly becomes enormously complicated with respect to the equal opportunity offered by those "culture-free" tests.

Finally, the telling blow against the term culture-free is its application to tests that we know very well to be influenced heavily by the totality of environmental experience that any person brings to an examining situation. Anastasi (1954, pp. 255–57; 1958, pp. 561–63) long ago pointed out the inappropriateness of "culture-free" and suggested "cross-cultural" as a replacement. Basically, no test is free of cultural influences, and "culture-free" should be dropped from further serious description of individual and group differences.

The waning popularity of one term however does not prevent another equally deceiving one from taking its place. Lately, we have witnessed the growing use of "culture-fair" as a substitute for culture-free. Many psychologists have merely started using the former label instead of the latter. Hence, a culture-fair test for them is one that demonstrates lesser differences among the means of different cultural (or ethnic) groups. In fact, at least one major test publisher has lately put on the market a test which, according to the examiner's manual, was designed to be fair for several cultural subgroups by reduction of intergroup differences in mean score. I hope my previous discussion of this approach has demonstrated its simplistic nature.

Although "culture-free" and "culture-fair" labels on tests are not likely to arouse many passions, the emotionally laden epithets of "discriminatory" and "culturally biased" are often applied to tests by people who ought to know that the mere demonstration of significant group differences in mean scores does not make a measuring instrument biased or illegally discriminatory. Charges that tests are inherently biased or discriminatory have no more rational basis than claims of their freedom from cultural influences or their fairness based on item content or superficial psychometric considerations.

Fortunately, there is an escape from the semantic jungle. Within the past year, two sets of authors (Kirkpatrick *et al.*, 1968; Bartlett & O'Leary, 1969) have tied the fairness of employment methods to their validity and to employee performance on the criterion, in addition to the mean predictor (e.g., test) scores. Somewhat earlier, Guion (1965, pp 491–93) had suggested the use of race as a moderator variable to determine whether tests are equally valid, hence fair, for minority and nonminority groups.

The scope of this paper does not allow a comprehensive discussion of the moderator variable approach to validation. In essence its goal is the improvement of precision in prediction by separating a heterogeneous group of persons into relatively more homogeneous groups and computing separate validity results for each group. Thus, an applicant group composed of Negroes and whites, both male and female, could be tested, hired, and later evaluated on their job performance. Assuming the exercise of proper controls on extraneous variables that might affect the criterion measures, the larger group could be moderated on both race and sex. In this example, four separate groups would be obtained—Negro males, Negro females, white males, white females. Needless to say, the subgroup sample sizes must be large enough to warrant this method of validity refinement.

The interested reader should consult the Kirkpatrick and Bartlett references and Katzell (1969) for some of the many pos-

sible outcomes of moderator variable technique as applied to equal employment opportunity. An article in the *Columbia Law Review* (1968) also demonstrates possible differential validation outcomes originally prepared by Richard S. Barrett. It should be noted that the use of moderator variables does not guarantee an increase of validity in any of the subgroups. A highly significant validity finding on a heterogeneous sample may be fragmented into two or more scatterplots with negligible relationship between the predictor and criterion when the differential validity concept is applied.

In addition to the results presented by Lopez (1966) on toll collectors, there is more recent evidence (Mitchell, Albright, & McMurry, 1968; Ruda & Albright, 1968) that test scores do not necessarily bear the same relation to job performance for Negroes as they do for whites. The Mitchell *et al.* study on male, hourly-paid, mostly semi-skilled workers showed Negroes making much lower test scores than whites; yet, their rated job performance and actual job tenure were not significantly different. The test had no significant validity for whites or Negroes against either criterion. Using workers in an office where turnover was a severe problem, Ruda & Albright found significantly lower test scores for the Negro sample than for the whites. The criterion was tenure, and Negroes tended to stay on the job longer than whites. Test validity for whites was significantly negative, while it was not significant for Negroes. Since the employer had been selecting both whites and Negroes with the higher test scores, whites were being chosen who were least likely to remain on the job. The test was obviously irrelevant in selection of Negroes.

These two studies in combination with the results obtained by Lopez and by Kirkpatrick *et al.* show the real need for much more differential validation research. And as the latter group of researchers point out,

> Unfair discrimination between ethnic groups cannot be inferred from evidence of differences in validity alone; *mean job criterion performance* (emphasis added) must also be considered. (p. 7.)

Thus, the way to make an employment system "fair" is to validate and use it in such a way that the probabilities of being hired or promoted vis-a-vis the probabilities of job success are the same for minorities and nonminorities. If expectancy charts indicate the same probability of job success for minority and nonminority groups, the minority group must enjoy the same proportion of new hires or promotions as the nonminority group. This must be done regardless of the personnel assessment method used, test or no test.

If the results of differential validation studies are applied in all seriousness to ensure equality in employment opportunity, occasions may arise when results indicate that minority applicants with lower test scores should be hired instead of other applicants with higher test scores. The justification for this action would be that the lower test score for a minority applicant predicts a higher criterion score than does the higher test score for a nonminority applicant. Many employers are squeamish about this action because they believe they may be guilty of "reverse discrimination" by using a double standard based on race, color, etc. My suggestion, as a pragmatic and scientifically defensible solution to this problem, is the conversion of *all* applicants' test scores to predicted criterion scores—using the appropriate regression equation of the moderator subgroup to which each applicant belongs. Predicted criterion scores should be ranked from highest to lowest, and selection can then be made from the top downward. There is no reverse discrimination in this approach. After all, do we not give tests to

predict what an employee will do on the job?

In closing, I should like to make the following observations. Fairness of personnel assessment lies in all the system variables. It cannot be attached to tests alone. Employment tests do not themselves discriminate against minority groups. People can; and some do. Tests do not screen out or screen in applicants for employment; people do. Tests do not exercise judgment or make personnel decisions; people do. Tests do not hire and promote; people do.

REFERENCES

Anastasi, A. *Psychological testing.* New York: Macmillan, 1954.

Anastasi, A. *Differential psychology* (3rd ed.). New York: Macmillan, 1958.

Bartlett, C. J. & O'Leary, B.S. A differential prediction model to moderate the effects of heterogeneous groups in personnel selection and classification. *Personnel Psychology,* 1969, *22,* 1–17.

Columbia Law Review Staff. Legal implications of the use of standardized ability tests in employment and education. *Columbia Law Review,* 1968, *68,* 691–744.

Equal Employment Opportunity Commission. *Guidelines on employment testing procedures.* Washington: EEOC, 1966.

French, R. L. The Motorola case. *The Industrial Psychologist,* 1965, *2,* 29–50.

Guion, R. M. *Personnel testing.* New York: McGraw-Hill, 1965.

Katzell, R. A. Statement from the Office of Federal Contract Compliance. In Chap. IV of *Now Hear This! Equal Employment Opportunity: Compliance and Affirmative Action.* New York: National Assn. of Manufacturers, 1969.

Kirkpatrick, J. J.; Ewen, R. B.; Barrett, R. S.; & Katzell, R. A. *Testing and fair employment.* New York: NYU Press, 1968.

Lopez, F. M., Jr. Current problems in test performance of job applicants (I). *Personnel Psychology,* 1966, *19,* 10–18.

Mitchell, M. D.; Albright, L. E.; & McMurry, F. D. Biracial validation of selection procedures in a large Southern plant. *Proceedings,* 76th Annual Convention, APA, 1968, 575–576.

Ruda, E. & Albright, L. E. Racial differences on selection instruments related to subsequent job performance. *Personnel Psychology,* 1968, *21,* 31–41.

Tiffin, J. & McCormick, E. J. *Industrial psychology* (4th ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1958.

CHAPTER V

Misuses of Tests

WILLIAM H. ENNEIS

Presented August 31, 1969, as part of a symposium, *Testing as a Social Problem: Issues and Responsibilities*, at the 77th Annual Convention of the American Psychological Association in Washington, D. C.

* * *

Many undergraduate college students are exposed to the fundamentals of psychological testing in a variety of courses offered to psychology, education, and business majors. These courses include elements of measurement theory, test construction, test administration, test scoring, and test validation—that is, all the technical niceties of our profession. This instruction is fundamental, but rarely does the undergraduate student get any instruction on the social implications of test use—nor for that matter did many masters or doctoral students taking courses in personnel testing become aware of these problems until only a few years ago (Guion, 1965).

Although we as psychologists have set up high standards for test use (APA, 1966), most of our students have few additional contacts with our profession after they graduate. Those students, by their very numbers, ultimately have more control over the equitable use of tests than we do. But today most of them remember only that test instructions should be read completely; that lighting, heat, and ventilation should be properly controlled; that standard time limits on tests should be strictly observed; that all examinees should be treated courteously; that pencils should be sharpened before the testing session; that smudges on completed papers should be erased; that tests and answer sheets should be passed systematically across rows for collection; etc.; etc.; etc.

In the field of industrial testing, most of the persons who effectively control use of tests, such as personnel managers and administrators, are ignorant of the power they wield by the use of employment tests. Many of them do not have the social perceptiveness necessary to understand and to remedy the misapplication of employment tests to which they have materially contributed within the past decade. Perhaps, in reality, the misuse of tests has grown no more rapidly than the introduction of new personnel testing systems; but public awareness of potential misuse has definitely risen as a result of social and economic action programs. The remainder of this paper is devoted to some important misuses of personnel tests.

USE OF UNVALIDATED TESTS

This condition has existed since personnel testing escaped the control of the pioneer innovators in test construction and application and became a part of the trappings of the hack and the charlatan. I do not know when this really became a problem. Its history would make a good contribution to the understanding of our field. Validation of personnel tests for their intended purposes has been recommended by

professionals in the field as long as texts have been written on the subject. The fact exists, however, that most employers do not bother to determine whether test scores are systematically related to employee performance.

The failure to establish criterion-related validity of tests has two serious potential consequences: denial of employment to minorities and a waste of the employer's money.

Minority Employment

The fact that minorities make lower average scores on many tests is no longer a disputed fact. The conclusion that they do less well in the entire spectrum of jobs because of these lower test scores is not accordingly documented. Indeed, much evidence has been accumulated that minorities' test scores may underestimate their job performance if a single cutoff score, based only on the norms of the majority or a mixed ethnic group, is used in applicant selection where minorities constitute an identifiable factor in the labor force (Lopez, 1966; Kirkpatrick, Owen, Barrett, & Katzell, 1968; Bart & O'Leary, 1969).

In the absence of validation evidence, an employer may be using a test whose only known function is rejection of minority applicants (or employees) in greater proportions than nonminorities. Disproportionately high minority rejection from employment is a serious social problem by itself; and when no corresponding, useful business function of the test use has been demonstrated, many legal problems arise as well (Columbia Law Review Staff, 1968; Cooper & Sobol, 1969).

Waste of Employer's Money

Explicit discussion of the personnel administrator's responsibility to his employer for justification of funds expended on selection programs is notably lacking from many standard texts in the field. Methods for presenting to management the results of test validation are covered by some authors (e.g., Thorndike, 1949), but they are oriented more toward technique than responsibility. Psychologists and personnel administrators too often rely on their "professional judgment" or "expertise" to sell the idea of testing programs to management. Some consultants never insist that their clients conduct validity studies as a part of the service they provide to industry, and many resident personnel managers cling to unvalidated tests with a zeal that approaches fanaticism.

As a result of these circumstances, most corporate officials do not know whether their firm's personnel testing programs produce a financial return on their annual expenditures. It is commonly accepted in business that funds should not only be recovered by the activities or items for which they are allocated but that there should be a reasonable return on whatever investment is incurred. This principle is commonly accepted and demanded in the production, sales, advertising, and research and development of services and goods. For example, a production engineer who recommends installation of a new, operational assembly line will be asked to produce hard data to justify both initial capital investment and future operating and maintenance costs of the proposed system. If he could not do so and said that a salesman had told him the assembly line "worked" in other companies, with no accompanying justification to the present situation, I am sure that the engineer would—if he were not fired—be told that he must produce evidence of cost saving, either immediate or future, over the current assembly methods. This is not the case with the installation of most employ-

ment testing programs. They are usually installed uncritically without evidence that they will help to produce a more efficient work force. Indeed, many employers not only fail to achieve a return on the cost of their testing programs but also fail to recover even their basic expenditures because the tests have no validity for employee selection. In conclusion, I cannot agree that tests are "good, economically-sound selection procedures" (as I have seen them represented) unless they have been proved to be so within the context of hard-nosed business and professional standards.

USE OF NORMS IN LIEU OF VALIDITY DATA

The establishment of local test norms is indispensible to the administration of an effective employment testing program, *given the fact that the employer has validity evidence for his jobs.* However, the collection of normative data without validation is a waste of time. If test scores bear no systematic relationship to criteria of employee performance, then no raw-score cutoff, based on percentile ranks, will improve the quality of the employer's work force. Furthermore, establishment of test hiring standards without validation will almost invariably result in disparate rejection rates between minority and nonminority applicant groups. Of course, test validation does not guarantee equality of hiring rates, but the results have the salutary effect of putting into proper perspective the real function of a testing program.

IRRELEVANT TESTS AS MEASURES OF EMPLOYEE PERFORMANCE

Job skills tests, as measures of proficiency, have long been accepted by psychologists on the basis of their content relevance to the duties of specific jobs. For example, typing tests are frequently administered to persons who apply for work as typists and claim skill in that area. As along as such miniature job samples are known to produce stable results, the minimum qualifying scores are reasonably consonant with normal, expected productivity, and the actual job requires the skill in question to a significant degree, such job skills tests are reasonable requirements for initial hiring or criteria of proficiency following training.

However, tests that purport to measure training effectiveness or the proficiency of a fully trained employee can easily be misused as criteria against which general and special ability tests for employee selection are validated. First, the content of the test, as criterion, may represent an insignificant component of the total skills required for effective job performance. Second, the test-oriented measure of training or job performance may be designed in a way that the *form of response elicitation* is irrelevant to normal job requirements. This fault is exceedingly critical and may result in the false conclusion that the selection instrument is a good measure of employees' potential productivity. For example, consider as a measure of training proficiency a paper-and-pencil test that is basically constructed and administered like a general learning ability test. If the training proficiency test has a time limit, uses a complex item and response format, and measures vocabulary and reading comprehension as much as requisite job knowledge, it may spuriously validate a predictor test because both of them measure academic skills and "test-taking" ability rather than actual job fundamentals. Indeed, it is possible that neither the original employment test nor the proficiency test would be predictive of on-the-job performance in these circumstances.

The primary danger of using tests as

criteria is that a statistically valid but fundamentally irrelevant employee selection system might be established that favors job applicants with test-taking skills. Employee selection systems of that type cannot be defended on business, professional, or social grounds.

CONFUSION OF PRESENT ACHIEVEMENT LEVEL WITH LEARNING ABILITY

Wesman (1968) has neatly punctured the artificiality of distinctions in the classification of tests as those of ability, aptitude, and achievement. His main points are that all our educational and employment tests are *ability* tests, that all ability tests are achievement tests which tap the product of learning and biological structure, and that the only real difference between an achievement test and an aptitude test is the *purpose* for which it is given. I should like to amend that distinction by stating that achievement and aptitude tests should be categorized by their *known functions* rather than their intended use.

The assumption that present achievement will predict future achievement (of a different variety) is the logical basis for the use of ability tests as employment screening devices; and, strictly speaking, a test that fails to predict future achievement of some sort has no right to be called an *aptitude* test. The entire basis of the use of achievement tests as aptitude tests in employment is further predicated on the assumption that all applicants have been exposed to the same general opportunities for learning, since, on the basis of the "equal exposure" concept, those persons who have the greater capacities for learning will have achieved more, as measured by tests, and would be the more likely persons to learn, for example, job skills.

Given the past and present conditions of our society, the "equal exposure" (or "equal opportunity") principle for learning and achievement, as measured by employment tests, is completely false. When minority groups are tested for their potential as employees, their unequal opportunities for learning the content of employment tests are all too clear. Under these circumstances, the fact that disadvantaged groups have not learned the skills propaedeutic to test-taking does not mean that they are unable to learn job-related skills, including skills of a highly complex nature. Thus, so-called "aptitude" tests given to assess employability are often nothing more than indicators of past achievement —not future potential.

GENERAL ABILITY TESTS FOR SELECTION OF EXPERIENCED WORKERS AND COLLEGE GRADUATES

Employers still use general ability ("intelligence") tests to screen college graduates and persons with known job skills. This practice is closely related to the confusion of achievement with aptitude; but in this case it is even worse because the achievement, as demonstrated by college graduation or prior work experience, is a more relevant indicator of employee potential than a single test score.

A college graduate has already demonstrated the ability to learn. If the employer wants to garner further relevant information (in addition to grade-point average), he should use validated tests with content specific to the curriculum that the graduate offers as part of his credentials.

An experienced worker brings to the employer professed knowledge and skills. Both can be assessed by trade information and job skills tests.

Reliance on general ability tests to screen either graduates from accredited colleges or experienced workers is a symp-

tom of the laziness and incompetence that characterizes the personnel operations of many employers, especially those of the bureaucratic type whose rigidity and dependence on the statistical trappings of professionalism have led them into the quixotic search for perfectly internally consistent tests where error variance is zero and test reliability is plus one. My comment is that such employers had better start worrying less about the technicalities of error variance reduction and concern themselves more with an increase in true variance against meaningful criteria of employee performance.

ACCEPTANCE OF CONCURRENT VALIDITY AS A CONSERVATIVE ESTIMATE OF PREDICTIVE VALIDITY

Concurrent validation of tests is apparently the scourge of industrial psychology. I have never seen an educational psychology text that describes this method as acceptable for the prediction of students' future achievement; however, industrial psychologists frequently use it inside and outside of academic settings to determine the validity of tests for student and trainee success in courses and training programs that lead to vocations or specific jobs.

Concurrent validity studies are generally acknowledged to produce conservative estimates of predictive validity, primarily because curtailment of range in test or criterion scores, or both, militates against the demonstration of any "true" correlation that may exist between the test and the criterion. However, Ryan & Smith (1954, p. 71) point out that job training and experience may affect not only employees' work performance but also their scores on tests administered after they are productive workers. They continue by noting that a perfect correlation between amount of training and the test would raise considerably the test scores of those who have learned the job well. Under these conditions, test scores obtained from present employees would produce a spuriously inflated validity coefficient and lead to the erroneous conclusion that the test administered to applicants for employment would result in satisfactory prediction of their job performance.

The situation related here is certainly not the general rule. However, for the educationally disadvantaged, jobs in a firm with a good training program may represent their first opportunity to profit from systematic instruction. These conditions may significantly alter the validities of employment tests in ways that are generally unknown and that deserve considerable research.

INDISCRIMINATE TEST SALES

Although the APA (1966, pp. 10–11) has established levels of tests requiring different qualification standards of persons who administer them and interpret results, some test publishers seem not to have taken these recommendations seriously. Sales of all sorts of tests are being made to persons who have *no* psychological training whatsoever. Some of these instruments, sold as employment tests, are basically clinical devices that should never be used by the persons to whom they are sent. The more responsible test publishers have set up standards for test sales to various classes of users, but the extent to which test purchase orders are screened to determine eligibility of the buyer is another matter. The obligations of publishers and distributors in policing the use of their tests are extremely complex, and there is no easy answer. However, with the recent generation of many new ideas concerning the social responsibilities of

business enterprises, the test publishing industry should begin to examine more closely the practices of test users whose efforts toward professional application of test results are lacking.

CONCLUDING REMARKS

Although this paper has dealt with the misuses of psychological tests, many things I have said are applicable to other employment procedures as well. Tests, like other employment methods, are neither good nor bad away from the contexts in which they are used.

It is often said that tests are "discriminatory" and "culturally-biased." This simply is not true. Employment tests do not discriminate against minority groups. People can; and some do. Tests do not screen out or screen in applicants for employment; people do. Tests do not exercise judgment or make personnel decisions; people do. Tests do not hire and promote; people do.

REFERENCES

American Psychological Association. *Standards for educational and psychological tests and manuals.* Washington: APA, 1966.

Bartlett, C. J. & O'Leary, B. S. A differential prediction model to moderate the effects of heterogeneous groups in personnel selection and classification. *Personnel Psychology,* 1969, *22,* 1–17.

Columbia Law Review Staff. Legal implications of the use of standardized ability tests in employment and education. *Columbia Law Review,* 1968, *68,* 691–744.

Cooper, G. & Sobol, R. B. Seniority and testing under fair employment laws: A general approach to objective criteria of hiring and promotion. *Harvard Law Review,* 1969, *82,* 1598–1679.

Guion, R. M. *Personnel testing.* New York: McGraw-Hill, 1965.

Kirkpatrick, J. J.; Ewen, R. B.; Barrett, R. S.; & Katzell, R. A. *Testing and fair employment: Fairness and validity of personnel tests for different ethnic groups.* New York: NYU Press, 1968.

Lopez, F. M., Jr. Current problems in test performance of job applicants (I). *Personnel Psychology,* 1966, *19,* 10–18.

Ryan, T. R. & Smith, P. C. *Principles of industrial psychology.* New York: Ronald Press, 1954.

Thorndike, R. L. *Personnel selection: Test and measurement techniques.* New York: Wiley, 1949.

Wesman, A. G. Intelligent testing. *American Psychologist,* 1968, *23,* 267–274.

CHAPTER VI

Minority Employment Barriers From the EEOC Viewpoint[1]

WILLIAM H. ENNEIS

Presented September 2, 1969, as part of a symposium, *The Black Man in the World of Work,* at the 77th Annual Convention of the American Psychological Association in Washington, D. C.

* * *

INTRODUCTION

It would be easy to say that unemployment and underemployment of minorities are mostly the product of discrimination and let it go at that. Of course, the use of "discrimination" in that sense is as meaningless as the explanation of high accident rates among some workers on the basis of their "accident-proneness" and the ascription of wars to "human nature" or an "instinct" for aggression. The purpose of this paper is an explanation of fallacies that permeate the thinking and the practices of persons who establish the conditions under which employment decisions are made, because it is the adverse effects of these fallacies that constitute significant barriers to the employment and advancement of minorities in the world of work. The last section explores the concept of system validation.

CONFUSION OF SELECTION STANDARDS WITH JOB REQUIREMENTS

Many employers and unions insist that they cannot "lower their standards" and use this argument as a defense for low utilization of minorities. However, these standards are nearly always defined in terms of information gathered during the applicant screening process. Only the rare employer or union can show that its standards for hiring (or later promotion) are significantly and meaningfully related to the requirements of work performance among its employees or members. In fact, most employers and unions cannot even state in objective or logically consistent form the amount and quality of work that are expected from their employees or members for different kinds of jobs.

What are some of these pre-employment standards that employers use to screen applicants? They include test scores, years of schooling, illegitimate children (for women but *not* for men), casual and often highly subjective reactions of interviewers, and police records. Recently, we encountered a situation in which an apparently otherwise qualified young Negro was rejected as a stockbroker trainee because his "social intelligence" was not high enough, based on a psychological instrument which, by its title, purports to measure that quality. This black college graduate had passed all the other employment tests, no other reason was given by the securities firm for his rejection, and the company had no validity evidence whatever to support use of the social intelligence scale for the job in question.

[1]A shortened version of this paper appeared in *Professional Psychology,* I, (Fall, 1970), 435-439.

Obviously, the problem centers on the fact that people responsible for setting hiring qualifications have usually failed to validate their selection systems. By some mysterious process, the power to assess future work performance is transferred in the mind of the personnel administrator to selection instruments and methods for which no professional or business justification has been established.

The EEOC has never advocated that an employer lower requirements of productivity among members of his work force. However, the Commission has consistently urged that hiring standards or qualifications be systematically validated against employee job performance and, in some cases, has insisted that applicant screening methods and cutoff scores be altered when these selection methods have no demonstrated validity for the employer's jobs and also result in disproportionately high rejection rates among minority applicants or present employees. Only in this way can the employer establish that his selection procedures serve a real business need and that the cutoff level established is one below which a significantly greater proportion of applicants ultimately fail to meet standards of productivity normally expected of experienced employees. After all, if scores derived from a screening procedure are not significantly related to employee performance, no cutting point can be established that will result in a better work force, as defined by the criterion measure(s) used to assess productivity.

FAILURE TO CONSIDER DIFFERENTIAL VALIDITY

If relatively few employers validate their selection systems at all, far less have validated them separately on the different ethnic groups that comprise significant factors among the labor force from which their potential employees could be drawn. The concept of using ethnic groups of applicants (or employees) as a moderator variable in validation of predictors, especially tests, has received an increasing amount of attention during the past few years.

Krug (1966) published one of the first analytic models showing the need for validation of psychological tests on separate ethnic groups rather than large heterogeneous ones. A more recent analytic paper (Bartlett & O'Leary, 1969) includes summary results from validation studies in education and industrial settings and definitely shows that differential validity is a phenomenon that cannot be ignored. Lopez (1966) presented data on separate groups of Negro and white toll collectors which indicated that tests may not always predict job performance in the same way for different racial groups. Kirkpatrick, Ewen, Barrett, & Katzell (1968) also found that tests scores are not necessarily related to criterion performance in the same patterns for minority and nonminority groups. A study by Tenopyr (1967) on machine-shop trainees has been cited (e.g., Ruch, 1969) as showing that tests have the same validity for both Negroes and Anglos. However, a careful examination of her study reveals that only regression slopes were equal for both ethnic groups, while all test and many criterion means were significantly different. The concept of differential validity does not, as some people erroneously assume, concern itself solely with differences in validity coefficients (or regression slopes). This concept includes an analysis of differences among predictor and criterion means as well.

If significant differences are found in regression coefficients, predictor means, or criterion means among different ethnic groups for the same job(s), the application of different prediction equations becomes virtually mandatory, if from nothing more than a strictly professional

viewpoint. Whenever differential prediction is empirically justified, conversion of all predictor (e.g., test) scores to predicted criterion scores—using appropriate moderated regression equations—has been suggested as a way to eliminate apparent double standards and possible claims of "reverse discrimination" in applicant selection from various ethnic groups (Enneis, 1969).

Aside from purely professional issues, there are compelling social, ethical, and legal (Cooper & Sobol, 1969, pp. 1645–46, 1966–67) reasons for differential validation studies that use race or ethnic grouping as moderator variables. Black people and Spanish Surnamed Americans have long been categorically excluded from various types of work, especially in certain sections of the country. Although occupational segregation is illegal under Title VII of the Civil Rights Act of 1964, it continues de facto because selection procedures are generally geared to the white, Anglo majority. Since virtually all expectancy tables of predicted job success are based on all-white or mixed ethnic groups, use of a single regression equation may result in serious prediction errors for all applicants; and when minority groups earn, for example, lower average test scores but achieve success on the job equal to that of the majority group, use of the single regression equation based on the composite validation sample will result in higher, unfair rejection rates for minority applicants.

CHAIRMAN-OF-THE-BOARD SYNDROME

Shortly after I started working at the Commission, I coined this term to refer to application selection based on levels of attributes claimed necessary for the top job(s) in promotional sequences or lines of progression. The "chairman-of-the-board" syndrome reflects an idea that every person hired must ultimately be able to perform the most complex work in the promotional ladder. It is nearly always based on educational and psychological test standards, and it rarely takes into account that only a small fraction of those employees originally hired at entry levels reach the top jobs or that, if they do, the time period between hiring and attainment of the top job is normally several years.

This practice is particularly severe in predominantly blue-collar, unionized, manufacturing industries. It occurs at its worst among employers who have negotiated labor contracts which specify promotion based on seniority. The COB syndrome is one of the most pernicious in the entire American industrial system, and it often amounts to nothing more than an employee caste structure based on a cult of test score worship.

This early, one-shot evaluation of applicant potential might be partially justified if the initial hiring methods were valid for nearly all the jobs in the lines of progression. However, the unskilled, entry jobs do not require the arbitrarily set levels of education, general learning ability (achievement?), and mechanical aptitude (again, achievement?) that characterize the syndrome. Therefore, validation—when it has been done at ail—nearly always produces negative results. Even at the middle-level, semiskilled positions, significant validity findings are not common. At the top-level jobs, the educational and test requirements do sometimes demonstrate validity.

However, the fact that validity of hiring procedures and scores does not manifest itself until many years later has a tragic impact on minorities. First, this system denies them gainful employment in what are usually very good-paying jobs in the community, even at the unskilled and semiskilled levels. Second, no consideration is

given to the fact that persons with lower levels of education and lower test scores can often learn to perform all the duties of all the jobs in the lines of progression. The fact that minority applicants earn lower test scores or have less education, on the average, does not mean they cannot learn these jobs, especially when the waiting periods in most seniority systems are so long at each level and low test scores frequently represent lack of prior opportunity rather than inability to learn.

CONFUSION OF STATISTICAL VALIDITY WITH RELEVANCE

This issue is a far more subtle one than others discussed here. Nevertheless, it assumes importance in the present rush of some employers to validate their tests for equal employment opportunity compliance purposes.

Most students of organizational psychology would probably agree that criteria of employee performance should reflect fundamental goals of the employer. In other words, employee appraisal systems should form the core of an information system by which management can determine the success of not only its current operations but also its long-range goals. Likewise, applicant selection procedures should be an outgrowth of logical organizational planning and criterion development, not vice versa.

Currently, some employers are seizing on a multitude of criteria, at least one or two of which they hope will show significant relationships with tests that were instituted without any professionally accepted evidence of their possible relevance to job requirements—such as is derived from job analysis or, better, preplanned job functions and structure. Whenever applicant screening procedures are established before employee performance appraisal systems, arbitrarily chosen criteria that "validate" the predictor(s), so to speak, may represent nothing more than job-irrelevant cultural loading common to both predictors and criteria. Of course, a statistically significant relationship between predictor and criterion might represent fortuitous, logical relevance between the two, but a posteriori relationships of this sort do not build or advance strong foundations of a science. In addition, this situation may do nothing more than create and perpetuate, in the statistical sense, predictors that correlate well with superficial criteria of social acceptability but not with business requirements.

RESTRICTED RECRUITING

Historically, discriminatory recruitment was viewed in terms of overt exclusionary acts or statements of preferences for employees of a certain race, color, ethnic origin, religion, or sex. Sometimes, the "preference" was negatively phrased, such as, "No Colored," in classified job advertisements.

Recognition that the replenishment of a work need not be expressed so blatantly and yet have an exclusionary effect on minorities is a relatively new development. In fact, some aspects of traditional recruiting methods such as newspaper advertising, walk-ins, employee referrals, and failure to recruit among minority groups have attracted much attention as possible covert violations of equal employment opportunity laws and regulations (Blumrosen, 1968). For example, employee referrals, used exclusively as a recruitment source, will tend to perpetuate an existing all-white work force, given other patterns of social segregation that militate against informal communication of job opportunities to potential minority applicants. Also, reliance on walk-ins will generally lower the application and, therefore,

the employment rates of minorities when the employer is located in suburban areas near central cities with high minority concentrations.

Among industrial psychologists, this subject sparks little professional interest. Certainly the paucity of current research on recruiting methods testifies to that. Many professionals consider recruiting as a grubby chore. However, inasmuch as we do a great deal of research on selection, training, employee evaluation, systems and equipment design, leadership, management effectiveness, and organizational function and structure, we ought to devote more time to finding ways to get all people with the best potential to the right jobs. At least one minority employment barrier would fall by the wayside if we could do so.

SYSTEM VALIDATION

In the psychological literature, we always read that certain tests, interview items, biographical data, physiological responses, medical data, or training methods have served as predictor variables in a validation study for a certain type of job. Even when multiple or partial regression analyses are used, only a small part of the organizational context is included. The emphasis centers on the predictors in a validation study, and few other job or worker variables are measured and taken into account. Diligent investigators attempt to eliminate or statistically adjust for criterion contamination factors, but even those precautions cannot be assumed uncritically in reading the usual research report.

Because so many factors enter into the determination of predictor validity, it is not surprising to find that the average coefficient of correlation between tests and proficiency (job performance) criteria is approximately 0.20, while the average coefficient between tests and training criteria is about 0.30 (Ghiselli, 1966, p. 125). One can only speculate how much these averages may be inflated by the fact that nonsignificant findings are probably reported and published less frequently than significant results.

Rundquist (1969) discusses the "prediction ceiling" in personnel selection and concludes that both the validity ceiling of about 0.50 and the lack of a time trend toward improvement can be laid to development of predictors without consideration of man as an "open system" or the work environment as a complex system with highly specific components. In this context, he cites a series of studies by Ford and Meyer (see Rundquist, 1966) in which the original validity of a mathematical aptitude test for learning computer programming was reduced from more than 0.50 to near zero by instituting a special training program for the low aptitude trainees.

If a change in test validity this drastic can be effected by alteration of one system component, what is happening to lower or raise validity coefficients in industrial organizations where dozens of variables act in unknown and uncontrolled fashion? Frankly, we do not know. When one considers all the possible interactions of organizational variables, it should be no surprise that validities of tests are so highly erratic and so specific to the contexts in which they are used. Essentially we have been validating entire work systems as much as we have been validating tests.

The implications for minority employment opportunities are clear. With rapid changes in technology, development of new training programs (especially for the disadvantaged), increasing educational levels of minorities, more equitable recruiting systems, and job restructuring, no employer can safely assume that his selection instruments will maintain too

long whatever previous validity has been established for them. In view of the fact that the ethnic group of applicants and employees is a major organizational variable for most employers, we must do more research in the direction of differential validation. If we procrastinate and rest on past achievements in the field of personnel selection, we may soon jeopardize the reputation of industrial psychology among the scientific community and the general public.

REFERENCES

Bartlett, C. J. & O'Leary, B. S. A differential prediction model to moderate the effects of heterogeneous groups in personnel selection and classification. *Personnel Psychology*, 1969, *22*, 1–17.

Blumrosen, A. W. The duty of fair recruitment under the Civil Rights Act of 1964. *Rutgers Law Review*, 1968, *22*, 465–536.

Cooper, G. & Sobol, R. B. Seniority and testing under fair employment laws: A general approach to objective criteria of hiring and promotion. *Harvard Law Review*, 1969, *82*, 1598–1679.

Enneis, W. H. Personnel testing and equal employment opportunity. Paper presented at the meeting of the Pennsylvania Psychological Association, Mt. Pocono, Penna., June 1969.

Ghiselli, E. E. *The validity of occupational aptitude tests.* New York: Wiley, 1966.

Kirkpatrick, J. J.; Ewen, R. B.; Barrett, R. S.; & Katzell, R. A. *Testing and fair employment: Fairness and validity of personnel tests for different ethnic groups.* New York: NYU Press, 1968.

Krug, R. E. Some suggested approaches for test development and measurement. *Personnel Psychology*, 1966, *19*, 24–35.

Lopez, F. M., Jr. Current problems in test performance of job applicants (I). *Personnel Psychology*, 1966, *19*, 10–18.

Ruch, F. L. In Comments on psychological testing. *Columbia Law Review*, 1969, *69*, 608–618.

Rundquist, E. A. The prediction ceiling. *Personnel Psychology*, 1969, *22*, 109–116.

Tenopyr, M. L. Race and socioeconomic status as moderators in predicting machine-shop training success. Paper presented at the meeting of the American Psychological Association, Washington, D.C., September, 1967.

CHAPTER VII

Statement on Personnel Testing and Selection

WILLIAM H. ENNEIS

Presented at public hearings held June 2–4, 1970, in Houston, Texas, by the U.S. Equal Employment Opportunity Commission.

* * *

It is an old American belief that a person should be hired on the basis of his or her ability to do a particular job. Few people oppose that idea, and many endorse it —even though favoritism manifesting itself in a variety of forms often mocks the ideal.

A basic problem, however, is that existing employment methods and standards are rarely known to produce a better work force than might be obtained by other techniques. Only a small fraction of employers rigorously apply business principles to the operation of their personnel selection programs.

Thus, most top corporate officials do not know whether their firm's personnel practices, including those related to psychological testing, produce a financial return on their annual expenditures. It is commonly accepted in business circles that funds should not only be recovered by the activities or items for which they are allocated but that there should be a reasonable profit on whatever investment is incurred. The application of this principle is commonly demanded in the production, sales, advertising, and research and development of services and goods. This is not the case with most employment testing programs. They are frequently installed uncritically without evidence that they will help produce a more efficient work force. Indeed, many employers not only fail to achieve a return on the cost of their testing programs but also fail to recover even their basic expenditures because the tests have no validity for employee selection. Therefore, it cannot be argued that tests are "good, economically-sound selection procedures" (as they have been represented) unless they have been proved to be so within the context of hard-nosed business and professional standards.

During the past decade, there has been a notable increase in testing procedures of doubtful utility. Some companies in Houston have even installed elaborate and expensive personality and temperament inventories for routine production jobs in the face of repeated industrial research that shows them to be completely useless for most employee selection but just dandy for psychological Peeping Toms and the personnel office that wants to reject an applicant on any phony pretense of an ostensibly "objective" nature. An official of the Atomic Energy Commission has said that ". . . the artificial, non-job-related entrance requirement hides more bigotry than all the white pointed hoods in the country," and he suggested that our nation would have never developed into the world

power it is now if some present-day psychological testing standards of acceptability had been applied to screen persons who settled here (Herrick, 1968).

Tests though are not the sole employment hurdle. Educational standards—notably, demands for a high school diploma—are often set far higher than indicated as necessary by job analyses. In its recent *Guidelines on Discrimination Because of National Origin*, the Equal Employment Opportunity Commission said that it will "examine with particular concern" situations involving testing of English language skills and height and weight standards for employment where they are not required for the work to be performed (*Federal Register*, 1970). In these respects the *National Origin Guidelines* are quite similar to the Commission's earlier *Guidelines on Employment Testing Procedures*, issued August 24, 1966, in which a professionally developed ability test was interpreted as

> . . . a test which fairly measures the knowledge or skills required by the particular job or class of jobs which the applicant seeks, or which fairly affords the employer a chance to measure the applicant's ability to perform a particular job or class of jobs.

The confusion of standards of personnel selection and promotion with standards of employees' job performance has a catastrophic effect on the employment opportunities of minorities and women. The structure and content of contemporary recruiting and applicant evaluation methods result in disproportionately high rejection rates among these groups, usually without any supporting evidence of their business necessity. In the absence of validity evidence, an employer may be using a screening procedure whose only known function is rejection of minorities and women in greater proportions than non-minorities and men. Disproportionately high rejection of minorities is a serious social problem by itself; and when no useful business function of the employment procedure has been demonstrated, there are many Title VII problems as well.

The vast majority of employment tests in use today are measures of achievement, usually those of an academic nature. The assumption that present achievement will predict future job performance is the basic premise for the use of most employment tests. Furthermore, the use of such achievement tests as potential predictors of job performance is based on the additional assumption that applicants have been exposed to the same general opportunities for learning, since, on the basis of the "equal exposure" concept, those persons who have the greater capacities for learning will have achieved more, as measured by tests, and may be the persons more likely to learn, for example, job skills.

Given the past and present conditions of our educational systems, this "equal exposure" or "equal opportunity" principle for learning and achieving, as measured by most employment tests, is completely false. Under these circumstances, the fact that a large segment of minority groups have not learned test-taking skills does not mean that they are unable to learn job-related skills, including those of a highly complex nature. Thus, so-called "aptitude" tests given to assess employability are often nothing more than indicators of previous opportunity to learn—not future job potential.

On the other hand, many employers say, "We cannot lower our standards," in defense of their continued low utilization of minorities. This argument is particularly frequent among employers who have experienced significant technological change. There is a widespread notion that internally complicated and sophisticated equipment must be operated by the most intelligent persons available. This belief has yet to be universally proven; and, in fact,

there is quite a bit of evidence to show that as manufacturing processes become more and more automated, general intellectual requirements actually decline. Has anyone ever claimed that it requires greater intelligence, learning ability, or mechanical aptitude to operate a zipper than to button up a coat? Certainly not. And everyone knows that a zipper is far more intricate than a simple button and buttonhole.

The Equal Employment Opportunity Commission has never advocated that an employer lower productivity standards among members of his work force. However, the Commission has consistently urged that hiring standards or qualifications be systematically validated against employee job performance and has often insisted that applicant screening methods and test cutoff scores be changed when these selection methods result in disproportionately high rejection rates among minority applicants or present employees and have no demonstrated validity for the employer's jobs. Only in this way can the employer establish that his selection procedures serve a real business need and that the qualifying level established for hiring or promotion is one below which a greater proportion of applicants ultimately fail to meet standards of productivity normally expected from experienced employees. After all, if scores derived from a screening procedure are not related to employees' performance, absolutely no level of qualification for employment can be set that will result in a better work force, as determined by relevant measures of employee productivity and effectiveness.

Thus, it is high time that employers, unions, and employment agencies stop confusing tests, education, interviews, and application blanks with job requirements when they think of the "qualified" employee. Standards of employee performance are derived from job requirements and duties. They do not reside in test scores, years of schooling, and data from application blanks and interviews. In this respect, psychological tests have been highly touted by their publishers and their users on the basis of "objectivity" and freedom from the bias or prejudice which can operate, for example, during an interview. This is a highly specious argument because, from an equal employment opportunity viewpoint, no test is objective unless results from it are known to be directly related to measures of employee effectiveness for a particular job or class of jobs.

Perhaps too much attention has been directed to employment tests as "discriminatory" and "culturally-biased" instruments. The attack on tests has tended to obscure the fact that *it is people, not tests,* that practice employment discrimination. People can discriminate; and some do. Tests do not screen out or screen in applicants for employment; people do. Tests do not exercise judgment or make personnel decisions; people do. Tests do not hire and promote; people do.

In conclusion, irrelevant and unreasonable standards for job applicants and upgrading of employees pose serious threats to our social and economic system. The results will be denial of employment to qualified and trainable minorities and women, creation of disillusionment and frustration, spiraling labor costs, and erection of job barriers that are incompatible with both the necessities of American industry in particular and the goals of American society in general. The Commission will not stand idle in the face of this challenge. It will fight employment discrimination in whatever form it occurs. The cult of credentialism is one of our targets.

REFERENCES

Federal Register, Vol. 35, No. 8, p. 421, January 13, 1970.

Herrick, H. T. *Civil rights, gradualism, and the established order of things.* Speech at the AEC Industrial Relations Conference, Kansas City, Missouri, October 4, 1967.

CHAPTER VIII

Use of Nontest Variables in the Government Employment Setting

WILLIAM H. ENNEIS

Presented September 8, 1970, as part of a symposium, *Use of Nontest Variables in Admission, Selection, and Classification Operations,* at the 78th Annual Convention of the American Psychological Association, in Miami Beach, Florida.

* * *

INTRODUCTION

This paper contains no new results in the areas of personnel selection and classification. It offers no ready solutions to either old or contemporary problems. But it does recapitulate some earlier questions about the validity and utility of nontest variables, as well as raising a new one about the ethics of using immutable personal characteristics as standards of employability in any kind of job.

First, nontest variables in the Federal employment setting are covered, and as part of this discussion the concept of recruiting method as a selection device is considered. Second, questions about general research approach and special problems of applications of findings are posed.

FEDERAL NONTEST VARIABLES

The distinction between formal qualifying requirements and informal use of background data is an important one. Formal nontest job standards are established by the United States Civil Service Commission (CSC), and they essentially consist of: (1) type and amount of education or training and (2) prior job duties (especially the ones immediately preceding the position for which application is being made) and length of service in performance of those duties.[1] At times, special requirements, such as foreign language proficiency, are imposed; but these are rare in comparison to the total number of jobs filled.

If the applicant is already in the Federal service, the grade level and classification of his present job virtually dictate the next level to which he can be hired or promoted, although the job series classification becomes less important as the grade (salary) level increases. On the other hand, it must be noted that persons who have been trained or who have worked in a specific job title or have had extensive education in a particular field may qualify for several types of jobs, though perhaps at different grade levels. Thus, some psychologists can qualify as statisticians or as managerial candidates if they have the prerequisite education and/or job experience.

Whether such formal requirements pre-

[1] No psychological test results, such as those from the Federal Service Entrance Examination or special job skills tests, are considered here, even though such results often determine employment eligibility. Also, veterans' preference, a formal nontest variable, will not be discussed since military service is not a requirement per se, and credit given is a type of "bonus."

dict employee success is a question that will not be answered here. Recent publications by Berg (1970) and by Diamond and Bedrosian (1970) at least cast some doubts on the relationship between amount of education and quality of work performance *within* any given job classification. Amount of experience in one's present job may be a valid predictor of future performance; but the degree of such validity and the extent to which this validity has been used to establish qualifying limits for Federal employment have not been disseminated by traditional professional means to either industrial, psychological, or Federal employment circles. The real facts, if they are even known, are probably far more complex than the arbitrary standards prescribed to and used by Federal civil service classifiers.

The informal use of background or personal history data is an entirely different matter. In this situation, individual judgment—be it whim or perspicacious insight—rules absolutely.

A Federal official responsible for staffing his operations is confronted with many personal history items on the Personal Qualifications Statement (Form 171) that is submitted by each applicant. For some jobs, some of these items may be positively associated with one or more aspects of work performance. For other jobs there may be no validity for any of these items, and in other situations one should not discount the possibility of an inverse relationship between some facets of personal history and work criteria.

From my own personal experience and observation, Federal managers often give heavy weights to the following types of data: age, sex, marital status, eligibility status on a current CSC register, salary required, willingness to travel, availability date, name (or location) of high school attended, name (if any) of college attended, chief college subjects and major field(s) of study, honors or awards received, specific foreign languages in which proficiency is claimed, membership in professional and scientific groups, tenure on previous jobs, title of applicant's position in earlier jobs, name of previous employer, name of supervisor (especially from the prestige or personal acquaintance viewpoint), reason for wanting to leave present job, literary style and vocabulary used in description of prior job duties, names and occupations of references, prior military experience (aside from formal veterans' preference, some Federal managers have strong pro or con feelings about hiring former career military personnel), and any number of "personal suitability" items (health, convictions, etc.). Obviously, the information supplied may be of absolute importance. For example, an applicant who will not be available for six months cannot be considered for a position that must be filled within sixty days. However, this is an administrative necessity consideration—not a question of validity.

RECRUITING METHODS

An organization's recruiting procedure is a type of employee selection that is literally imposed on the total pool of job applicants. Persons outside the formal or the casual information networks "fail" to meet the job qualification standard of "knowing about" openings for which they may hold the necessary credentials. Within the Federal government, many jobs are "passed around" among people who have high access to these information networks. In one sense, a Federal manager can hardly be blamed for hiring his friends and acquaintances already employed by the Government. He avoids the uncertainties and time delays involved in getting CSC certification of an outside candidate, since the eligibility of a current

Federal employee is virtually assured for any reasonable type of promotion or transfer.

However, the effects of such recruiting on work force quality, possibly through changes in the selection ratio and validity (if any) of formal selection standards, should be a matter of vital concern to all personnel psychologists. Although extensive research has been conducted to determine the effectiveness of various recruiting methods on producing applicants for various jobs, there are virtually no studies that include recruiting technique as the independent variable and validity of predictors as the dependent variable.

BASIC QUESTIONS

At this point it seems appropriate to recount some of the questions that have been raised by other writers on the subject of nontest variables. I also want to add a few comments of my own.

1. What validity do nontest variables have at all for selection and placement? Is this validity greater than, less than, or essentially equal to that obtained from psychological tests for specific jobs? To what extent is this validity generalizable to related jobs and applicant groups?

2. To what extent is this validity independent of that contributed by psychological tests? As a corollary, how much is the standard error of prediction reduced by addition of nontest to test variables, or vice versa?

3. At what point in the selection/placement process should nontest variables be introduced? Is a multiple regression or a multiple cutoff method better?

4. Do validities of nontest variables stand up under cross-validation in the same job context? That is, do the initial validities of nontest variables rely more or less heavily on unique variance in the predictor and the criterion than the validities of tests?

5. To what extent is there a logical confusion between nontest predictors and criteria of job success? This confusion or direct overlapping of predictors with criteria is a type of criterion contamination that is often ignored by personnel psychologists. An example consists of sales managers' rating of their salesmen's job performance, at least in part, on grooming and personal appearance when the same characteristic has been rated as part of the selection procedure.[2]

6. What effect do moderator variables such as race, sex, religion, or national origin have on validity and fairness of nontest variables? Selection on these factors, as such, is prohibited within the Federal government and is illegal for private employers, unions, and employment agencies. Therefore, if any prohibited bases for selection are themselves highly correlated with nontest (or psychological test) variables, differential validation on the basis of race, sex, religion, or national origin becomes virtually mandatory to determine fairness of the nontest variables.

7. What is the fakability of responses to nontest instruments? Extensive research indicates that "favorable" responses can be given, on the average, by applicants confronted with personality, attitude, and interest inventories. Is the same true for biographical inventories? Since hiring standards often conform to middle-class norms, can the more astute candidates alter their responses to personal history items to conform to what they think are the "better" answers?

8. What is the cost of developing and

[2] Some may argue that grooming is important in sales work. As a potential predictor, yes. As a criterion, no. Sales personnel are hired to sell effectively—not to look pretty or affect the pose of the all-American boy!

using nontest variables in comparison to other methods? Is the valid nontest predictor less expensive to research initially (which I seriously doubt), or does it show an equal or greater return on investment by the user than psychological tests with respect to validity and, therefore, reduction in personnel costs to the employer?

9. Finally, what ethical obligations does a personnel psychologist incur when he recommends, on the basis of confirmed validity evidence, that his employer adopt nontest selection and placement standards that are immutable—that cannot be changed regardless of individual effort or intervening circumstances? Does it make sense (assuming validity for such items) to give a job candidate minus or plus values on a background questionnaire simply because he had few (or many) books in his home as a child or had no father (or mother) in the home, even though other factors would point to good potential? This is not a frivolous question because many personnel departments use our instruments and score them with a frightening zeal. With many of them, failure of a job candidate to score acceptably, by even one or two points, is akin to degeneracy. Guilford (1954, p. 416) reports a study by Travers that showed success as an administrative scientist to be associated with a rural (vs. urban) upbringing and father's occupation of craftsman (vs. small business operator). These background characteristics are obviously immutable, and the findings did not make sense rationally until it was discovered that many of the urban-reared scientists were Jewish and that the performance ratings of their administrative duties contained a definite anti-Semitic basis.

CONCLUSION

If anyone has reached the conclusion that I am against the use of nontest factors in personnel selection and classification, he is dead wrong. However, nontest methods should be developed and applied with all the scientific rigor and professional ethics we can muster. I sincerely hope that personnel and educational psychologists will meet this challenge.

REFERENCES

Diamond, E. E., & Bedrosian, H. *Industry hiring requirements and the employment of disadvantaged groups.* Report submitted to the Manpower Administration, U.S. Department of Labor, by New York University School of Commerce, 1970.

Berg, I. *Education and jobs: the great training robbery.* New York: Praeger, 1970.

Guilford, J. P. *Psychometric methods,* 2nd ed. New York: McGraw-Hill, 1954.

☆U.S. GOVERNMENT PRINTING OFFICE: 1971 0-409-985

Office of Research

Equal Employment Opportunity Commission

1800 G Street, N.W. Washington, D.C. 20506